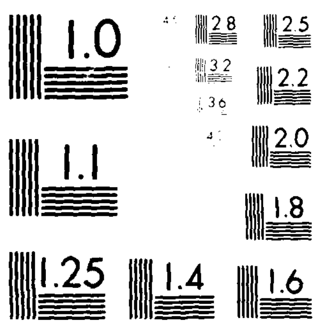
AD-A087 029 BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMI--ETC F/G 12/1
AVERAGING METHODS FOR THE ASYMPTOTIC ANALYSIS OF LEARNING AND A--ETC(U)
APR 80 H J KUSHNER, Y BAR-NESS, H HUANG N00014-76-C-0279

UNCLASSIFIED LCDS-TR-80-1 NL

1 OF 1
AD A
087029

END
DATE
FILMED
9-80
DTIC



MICROCOPY RESOLUTION TEST CHART

LEVEL II

LCDS TECHNICAL REPORT 80-1

I. AVERAGING METHODS FOR THE ASYMPTOTIC ANALYSIS OF LEARNING AND ADAPTIVE SYSTEMS, WITH SMALL ADJUSTMENT RATE

H.J. KUSHNER AND HAI HUANG

II. ANALYSIS OF NONLINEAR STOCHASTIC SYSTEMS WITH WIDE-BAND INPUTS

H.J. KUSHNER [redacted] Y. BAR-NESS

APRIL 1980

Lefschetz Center for Dynamical Systems

Division of Applied Mathematics

Brown University Providence RI 02912

AVERAGING METHODS FOR THE ASYMPTOTIC ANALYSIS OF LEARNING AND ADAPTIVE SYSTEMS, WITH SMALL ADJUSTMENT RATE

H. J. Kushner*

Hai Huang**

Abstract

Recently proved theorems concerning weak convergence of non-Markovian processes to diffusions, together with an averaging and a stability method, are applied to two (learning or adaptive) processes of current interest: (1) an automata model for route selection in telephone traffic routing, (2) an adaptive quantizer for use in the transmission of random signals in communication theory. The models are chosen because they are prototypes of a large class to which the methods can be applied. The technique of application of the basic theorems to such processes is developed. Suitably interpolated and normalized (learning or adaptive) processes converge weakly to a diffusion, as the (learning or adaptation) rate goes to zero. For small learning rate, the qualitative properties (e.g., asymptotic (large-time) variances and parametric dependence) of the processes can be determined from the properties of the limit.

---

*Divisions of Applied Mathematics and Engineering, Brown University, Providence, R.I. 02912. This research was supported by the Air Force Office of Scientific Research (AF-76-3063), the National Science Foundation (Eng. 77-12946), and the Office of Naval Research (N00014-76-C-0279-P003).

**Division of Applied Mathematics, Brown University, Providence, R.I. 02912. This research was supported by the National Science Foundation (Eng. 77-12946) and the Office of Naval Research (N00014-76-C-0279-P003).

I. INTRODUCTION

References [7], [1] develop a useful method to study the asymptotic properties as $\varepsilon \to 0$ and $n\varepsilon \leq T < \infty$ for any real T of solutions to stochastic difference equations of the form

$$(1.1) \qquad Y^{\varepsilon}_{n+1} = Y^{\varepsilon}_{n} + \varepsilon h_{\varepsilon}(Y^{\varepsilon}_{n},\xi^{\varepsilon}_{n}) + \sqrt{\varepsilon}\, g_{\varepsilon}(Y^{\varepsilon}_{n},\xi^{\varepsilon}_{n}) + o(\varepsilon), \qquad Y^{\varepsilon}_{n} \in R^{r},$$

where the distributions of the random sequence $\{\xi^{\varepsilon}_{n}\}$ might depend on the $\{Y^{\varepsilon}_{n}\}$. Such equations occur frequently in applications. The methods in [1] also work when ε is replaced by a sequence $\varepsilon_n \to 0$ as $n \to \infty$ from which asymptotic properties (rates of convergence) of various forms of stochastic approximations can be obtained.

The emphasis in [1] (an application of [7]) concerned the case where the h_{ε} and g_{ε} are smooth, and no details for the non-smooth case or its applications were given, nor was the asymptotic case where $n \to \infty$, then $\varepsilon \to 0$ treated. This is a deficiency, since in many applications in communication, control and automata theory, the h_{ε} and g_{ε} might simply be indicator functions and the noise $\{\xi_n\}$ depend on $\{Y^{\varepsilon}_{n}\}$, and the asymptotic properties (as $n \to \infty$, then $\varepsilon \to 0$) desired. Here, we apply the basic results of [7] to two such problems. The two problems have current technological importance in their own right and each has been the subject of a great deal of work. Our method often yields a complete analysis of the asymptotic properties under realistic conditions. The two problems are typical of a wide class, and they illustrate the power and applicability of the general technique, as well as the method of applying it to concrete problems. In a sense the method is an extension with more complex memory structure of the sort of "slow learning" results obtained by Norman [9], and should have broad applications to the areas cited above.

The basic type of result is the following. Define $Y^\varepsilon(\cdot)$, $t \in [0,\infty)$, by $Y^\varepsilon(0) = Y_0^\varepsilon$ and $Y^\varepsilon(t) = Y_i^\varepsilon$ on $[i\varepsilon, i\varepsilon+\varepsilon)$. Under appropriate conditions, Theorem 1 gives weak convergence of $\{Y^\varepsilon(\cdot)\}$ in $D^r[0,\infty)$ to a particular diffusion process, as $\varepsilon \to 0$. Now, let $\{n_\varepsilon\}$ denote a sequence of integers tending to ∞ as $\varepsilon \to 0$. For $t \geq 0$, define $\tilde{Y}^\varepsilon(t) = Y^\varepsilon(t+\varepsilon n_\varepsilon)$. The tilde $\sim$ always denotes a shift by n_ε (discrete parameter) or $\varepsilon n_\varepsilon$ (continuous parameter). By using Theorem 1 but starting $\{Y_n^\varepsilon\}$ at time n_ε instead of at time 0, we will get a great deal of information on the asymptotic properties (large n, small ε). The next section gives some background material from [7]. Sections III to VI treat a learning automata approach to certain problems in adaptive routing of telephone calls [2]-[3]. The second problem, in Sections VII-VIII, concerns the asymptotic theory of an adaptive quantizer from communications applications [4], [5].

II. SOME BACKGROUND MATERIAL

$D^r[0,\infty)$ denotes the space of R^r-valued functions on $[0,\infty)$ which are right-continuous and have left-hand limits, and is endowed with the Skorokhod topology [6]. $\hat{\mathscr{C}}_0$ denotes the continuous functions on $R^r \times [0,\infty)$ with compact support and $\hat{\mathscr{C}}_0^{\alpha,\beta}$ the subset whose mixed partial derivatives up to order α in t and β in the components of x are continuous. Let $b_i(\cdot,\cdot)$, $a_{ij}(\cdot,\cdot)$, $i,j \le r$, be continuous functions on $R^r \times [0,\infty)$. Let the operator

$$A = \sum_i b_i(x,t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j} a_{ij}(x,t) \frac{\partial^2}{\partial x_i \partial x_j}$$

be the infinitesimal operator of a diffusion process $X(\cdot)$. Assume that the solution to the martingale problem (on $D^r[0,\infty)$) of Strook and Varadhan [8] corresponding to A has <u>a unique non-explosive solution for each initial condition</u>.

Let $b_N(\cdot)$ denote a function with values in $[0,1]$, equal to 1 on $S_N = \{x: |x| \le N\}$, equal to zero in $R^r - S_{N+1}$ and with second derivatives bounded uniformly in x and N. Define $\{Y_n^{\varepsilon,N}, n \ge 0\}$ by

$$(2.1) \qquad Y_{n+1}^{\varepsilon,N} = Y_n^{\varepsilon,N} + [\varepsilon h_\varepsilon(Y_n^{\varepsilon,N},\xi_n^\varepsilon) + \sqrt{\varepsilon} g_\varepsilon(Y_n^{\varepsilon,N},\xi_n^\varepsilon) + o(\varepsilon)] b_N(Y_n^{\varepsilon,N}),$$

$$Y_0^{\varepsilon,N} = Y_0^\varepsilon \text{ if } |Y_0^\varepsilon| \le N \text{ and is zero otherwise,}$$

and define $Y^{\varepsilon,N}(\cdot)$ analogously to $Y^\varepsilon(\cdot)$. For purely technical reasons, it is convenient to state the theorem in terms of $\{Y_n^{\varepsilon,N}\}$. Let A^N be the infinitesimal operator of a (not necessarily unique) diffusion process, denoted by $X^N(\cdot)$, and suppose that its coefficients $a_N(\cdot,\cdot)$, $b_N(\cdot,\cdot)$ are continuous, bounded,

have compact support and equal $a(\cdot,\cdot)$, $b(\cdot,\cdot)$ in S_N. Suppose that $\{Y^{\varepsilon,N}(\cdot)\}$ converges weakly to some such $X^N(\cdot)$ as $\varepsilon \to 0$, for each N. Then [7] $\{Y^{\varepsilon}(\cdot)\}$ converges weakly to $X(\cdot)$ as $n \to \infty$. The following theorem is a restatement of Theorem 3 of [7] with $\tau_\varepsilon = \varepsilon$. Theorem 2 of [7] provides a very convenient method of proving tightness, and we will use it in the sequel. Let $E_n^{\varepsilon,N}$ denote expectation conditioned on $\{Y_j^{\varepsilon,N}, j \le n, \xi_j^{\varepsilon}, j<n\}$.

Theorem 1. Assume the conditions stated above on the solution to the martingale problem on $D^r[0,\infty)$ corresponding to operator A, and on A^N and $X^N(\cdot)$. For each N, and $f(\cdot,\cdot) \in \mathcal{D}$, a dense set (sup norm) in $\hat{\mathcal{L}}_0$, let there be a sequence $\{f^{\varepsilon,N}(\cdot)\}$ satisfying the following conditions: it is constant on each interval $[n\varepsilon, n\varepsilon+\varepsilon)$, at $n\varepsilon$ it is measurable with respect to the σ-algebra induced by $\{Y_j^{\varepsilon,N}, j \le n, \xi_j^{\varepsilon}, j<n\}$ and

$$(2.2) \qquad \sup_{n,\varepsilon} E|f^{\varepsilon,N}(n\varepsilon)| + \sup_{n,\varepsilon} \frac{1}{\varepsilon} E|E_n^{\varepsilon,N} f^{\varepsilon,N}(n\varepsilon+\varepsilon) - f^{\varepsilon,N}(n\varepsilon)| < \infty ,$$

and as $\varepsilon \to 0$ and for each t as $n\varepsilon \to t$,

$$(2.3) \qquad E|f^{\varepsilon,N}(n\varepsilon) - f(Y_n^{\varepsilon,N}, n\varepsilon)| \to 0,$$

$$(2.4) \qquad E\left|\frac{E_n^{\varepsilon,N} f^{\varepsilon,N}(n\varepsilon+\varepsilon) - f^{\varepsilon,N}(n\varepsilon)}{\varepsilon} - \left(\frac{\partial}{\partial t} + A^N\right) f(Y_n^{\varepsilon,N}, n\varepsilon)\right| \to 0.$$

Then, if $\{Y^{\varepsilon,N}(\cdot), \varepsilon_0 > \varepsilon > 0\}$ is tight in $D^r[0,\infty)$ for each N, where ε_0 does not depend on N and $Y^{\varepsilon}(0)$ converges weakly to $X(0)$, $\{Y^{\varepsilon}(\cdot)\}$ converges weakly to $X(\cdot)$, the unique solution to the martingale problem with initial condition $X(0)$.

III. AN AUTOMATA PROBLEM - INTRODUCTION

Narendra [2], [3] and others have studied the application of automata and learning theory to problems in the routing of telephone calls through a multi-node network and have suggested a variety of interesting automata models for this application. Under various assumptions (both explicit and implicit) they have stated convergence results in a number of cases. Generally, their results are applications of Norman's [9] results on slow learning. Here, we take one of their models and show how to apply Theorem 1 to get a much more complete asymptotic theory (large time) for small rate of change of the automata behavior (ε), and under more realistic conditions. The case dealt with here can readily be generalized - as will be commented on below. The example illustrates the power and usefulness of the approximation techniques used here. The algorithm should be considered as a prototype. It might not be the best, but it well serves to illustrate the method.

The problem formulation. Calls arrive at a transmitting or switching terminal at random at discrete time instants $n = 0,1,2,\ldots$, with P{one call arrives at nth instant} = μ, $\mu \in (0,1)$, P{>1 call arrives at nth instant} = 0. From the terminal, there are two possible routings to the destination, route 1 and route 2, the ith route having N_i independent lines - and can thus handle up to N_i calls simultaneously. Let $[n,n+1)$ denote the nth interval of time. The duration of each call is a random variable with a geometric distribution: P{call completed in the (n+1)st interval|uncompleted at end of nth interval, route i used} = λ_i, $\lambda_i \in (0,1)$. The members of the double sequence of the interarrival times and call durations are mutually independent. It is possible to work with more general Markovian arrival processes, but we retain a simple structure in order to emphasize the main points. In practice, a more complex

network would occur - and perhaps cycles might exist, and a vector routing parameter would be used, one component per node. But the main idea is similar. As in Theorem 4, the average dynamics are used for the stability analysis. From that point on, the proof of the appropriate generalization of Theorem 5 would be quite similar to the proof of Theorem 5.

The parameter ε will be used for the rate of adjustment of the routing automaton - the device which selects the route. The adjustment mechanism will be defined later. The routing automaton operates as follows. For each fixed ε, let $\{y_n^\varepsilon\}$ denote a sequence of random variables - with values in $[0,1]$. In order to have an unambiguous sequencing of events, suppose that the calls terminating in the nth interval actually terminate at time $n+\frac{1}{2}$, and arrivals and route assignments are at the instants 0,1,2,... precisely. <u>Thus the state at time</u> $(n+1)^-$ <u>does not include the calls just terminated or calls arriving at</u> $(n+1)$. Define the "route occupancy process" $X_n^\varepsilon = (X_n^{\varepsilon,1}, X_n^{\varepsilon,2})$, where $X_n^{\varepsilon,i}$ is the number of lines of route i occupied at time n^+. Thus, $X_n^{\varepsilon,i} \leq N_i$. If a call arrives at instant n+1, the automaton "flips a coin", and chooses route 1 with probability y_n^ε and chooses route 2 with probability $(1-y_n^\varepsilon)$. If all lines of the chosen route i are occupied at instant $(n+1)^-$, then the call is switched to route j $(j \neq i)$. If all lines of route j are also occupied at instant $(n+1)^-$, then the call is rejected, and disappears from the system.

In a more realistic situation, the network would have many nodes - not simply 2, and many possibilities of routing from node to node. The adjustment algorithm might be different, but the problem would be handled in exactly the same way. The object is to adjust the $\{y_n^\varepsilon\}$ sequentially (based on the system behavior) so that some desired behavior occurs. In order to be specific, we use the following "linear-reward" algorithm [3]. Let J_{in}^ε denote the indicator

of the event {call arrives at n+1, is assigned first to route i and is accepted by route i}. For practical as well as theoretical purposes, it is important to bound y_n^ε away from the points 0 and 1. Let $0 < y_\ell < y_u < 1$. We use the algorithm (3.1), where $\Big|_{y_\ell}^{y_u}$ denotes truncation at y_u or y_ℓ, and $\alpha(y) = 1-y$, $\beta(y) = -y$.

$$(3.1) \qquad y_{n+1}^\varepsilon = [y_n^\varepsilon + \varepsilon\alpha(y_n^\varepsilon)J_{1n}^\varepsilon + \varepsilon\beta(y_n^\varepsilon)J_{2n}^\varepsilon]\Big|_{y_\ell}^{y_u}.$$

Define $\alpha_\varepsilon(\cdot)$, $\beta_\varepsilon(\cdot)$ such that $\alpha(\cdot) = \alpha_\varepsilon(\cdot)$ in $[y_\ell, y_u-\varepsilon]$ and $\beta(\cdot) = \beta_\varepsilon(\cdot)$ in $[y_\ell+\varepsilon, y_u]$ and otherwise are such that (3.2) is equivalent to (3.1):

$$(3.2) \qquad y_{n+1}^\varepsilon = y_n^\varepsilon + \varepsilon[\alpha_\varepsilon(y_n^\varepsilon)J_{1n}^\varepsilon + \beta_\varepsilon(y_n^\varepsilon)J_{2n}^\varepsilon].$$

We will study the asymptotics of the behavior of a centered and normalized $\{y_n^\varepsilon\}$ for small ε. Part of the difficulty, which our scheme is well able to handle, is due to the fact that $\{y_n^\varepsilon\}$ is <u>not</u> Markovian. In the theoretical parts of [2], [3], the problem is set up so that $\{y_n^\varepsilon\}$ is Markovian.

<u>Some definitions</u>. If the choice probabilities y_n^ε are held fixed at some value y for all n, then the route choice automaton still makes sense, although there is no learning. For fixed route selection probability $y \in (0,1)$, let $\{x_n(y)\} = \{(x_n^1(y), x_n^2(y)), 0 \le n < \infty\}$ denote the corresponding route occupancy process. For the process $\{x_n(y)\}$, the state space $Z = \{(i,j): i \le N_1, j \le N_2\}$ (whose points are supposed ordered in some fixed way) is a single ergodic class, and the

probability transition matrix, denoted by $A'(y)$, has infinitely differentiable components. With given initial condition $\{P\{X_0(y)=\alpha\}, \alpha\in Z\}$ define $P_n(\alpha|y) = P\{X_n(y)=\alpha\}$ and the vector $P_n(y) = \{P_n(\alpha|y), \alpha\in Z\}$. Then

$$(3.3) \qquad P_{n+1}(y) = A(y)P_n(y).$$

The pair $\{(X_n^\varepsilon, y_n^\varepsilon), n\geq 0\}$ is a Markov process on $Z \times \{y_\ell, y_u\}$ and the <u>marginal transition probability</u> $P\{X_{n+1}^\varepsilon=(k,\ell)|X_n^\varepsilon=(i,j), y_n^\varepsilon\}$ is just the ((i,j)-column, (k,ℓ)-row) entry of $A'(y_n^\varepsilon)$. Define the vector $P_n^\varepsilon = \{P_n^\varepsilon(\alpha), \alpha\in Z\}$ where $P_n^\varepsilon(\alpha) = P\{X_n^\varepsilon=\alpha|y_\ell^\varepsilon, \ell<n, X_0^\varepsilon\}$. Then

$$(3.4) \qquad P_{n+1}^\varepsilon = A(y_n^\varepsilon)P_n^\varepsilon.$$

Also, let $P(y) = \{P(\alpha|y), \alpha\in Z\}$ denote the unique invariant measure for $\{X_n(y)\}$, with <u>marginal</u> defined by $P^1(j|y) = \sum_k P(j,k|y)$, $P^1(y) = \{P^1(j|y), j\leq N_1\}$, and similarly for route 2. Finally, define the transition probability $P(\alpha,j,\alpha_1|y) = P\{X_j(y)=\alpha_1|X_0(y)=\alpha\}$ and write the marginal as

$$P^i(\alpha,j,k|y) = P\{X_j^i(y)=k|X_0(y)=\alpha\}.$$

Define E_n^ε to be the expectation conditioned on $\{X_\ell^\varepsilon, y_\ell^\varepsilon, \ell\leq n\}$.

<u>A relationship of</u> (3.1) <u>to a differential equation</u>. Define $\nu_i = (1-\lambda_i)^{N_i}$. Note that

$$(3.5a) \qquad E_n^\varepsilon J_{1n}^\varepsilon = \mu y_n^\varepsilon[1 - \nu_1 I\{X_n^{\varepsilon,1}=N_1\}],$$

(3.5b) $$E_n^\varepsilon J_{2n}^\varepsilon = \mu(1-y_n^\varepsilon)[1 - \nu_2 I\{X_n^{\varepsilon,2}=N_2\}].$$

For small ε, it is reasonable to try to relate the behavior of $\{y_n^\varepsilon\}$ to the solution of (3.6), where $\hat{F}(y)$ is just $E[\alpha(y)J_{1n}^\varepsilon + \beta(y)J_{2n}^\varepsilon]$, but with $\{X_n^\varepsilon, y_n^\varepsilon\}$ replaced by $\{X_n(y), y\}$ and using the stationary measure.

(3.6) $$\begin{aligned}\dot{y} &= \mu\alpha(y)y[1-\nu_1 P^1(N_1|y)] - \mu(1-y)\beta(y)[1-\nu_2 P^2(N_2|y)] \\ &= \mu y(1-y)[\nu_2 P^2(N_2|y) - \nu_1 P^1(N_1|y)] \equiv \hat{F}(y).\end{aligned}$$

As y increases, $P^1(N_1|y)$ increases (and $P^2(N_2|y)$ decreases) monotonically. Thus, there is a unique point $\bar{y} \in (0,1)$ such that $\hat{F}(\bar{y}) = 0$. Also, $\hat{F}(y) > 0$ for $y < \bar{y}$ and $\hat{F}(y) < 0$ for $y > \bar{y}$. We assume that $\bar{y} \in (y_\ell, y_u)$ and we also make the apparently unrestrictive assumption that $\hat{F}_y(\bar{y}) \neq 0$. We actually will study the asymptotic properties of $U_n^\varepsilon \equiv (y_n^\varepsilon - \bar{y})/\sqrt{\varepsilon}$, for large n and small ε. In particular, let n_ε be a sequence of integers tending to ∞ as $\varepsilon \to 0$, and define the processes $\tilde{U}^\varepsilon(\cdot)$ by $\tilde{U}^\varepsilon(0) = U_{n_\varepsilon}^\varepsilon$ and $\tilde{U}^\varepsilon(t) = U_{n_\varepsilon+i}^\varepsilon$ on $[i\varepsilon, i\varepsilon+\varepsilon)$. When the $\tilde{U}^\varepsilon(\cdot)$ are dealt with, the $\{n_\varepsilon\}$ will either be explicitly defined or their values will be unimportant. We show weak convergence of $\{\tilde{U}^\varepsilon(\cdot)\}$ to the Gauss-Markov diffusion $u(\cdot)$ defined by (6.3). If $n_\varepsilon \to \infty$ fast enough as $\varepsilon \to 0$, then the limit $u(\cdot)$ is stationary. The general method can be applied to many other problems in learning, automata and systems theory.

IV. SOME PRELIMINARY RESULTS

In this section, we prove some auxiliary results concerning uniform convergence of $P_n(y)$ and its derivatives to $P(y)$ and its derivatives.

Theorem 2. For each $y \in [y_\ell, y_u]$, let $A'(y)$ denote a Markov transition matrix (continuous in y) such that the corresponding Markov chain $\{X_n(y)\}$ is ergodic with invariant measure $P(y)$. Then $P(\cdot)$ is also continuous and there is a $\delta > 0$ such that the eigenvalues of $A(y)$, except for the single eigenvalue unity, are bounded in absolute value by $1-\delta$ for all $y \in [y_\ell, y_u]$. $P_n(y)$ converges to $P(y)$ uniformly (and at a geometric rate) in $y \in [y_\ell, y_u]$ and in $P_0(y)$.

Proof. The last sentence follows from the penultimate sentence. The continuity of $P(\cdot)$ is a consequence of the uniqueness for each y, of the eigenvector of $A(y)$ corresponding to the eigenvalue unity (the invariant measure). Next, suppose that there is no such δ. Let $A(y)$ be a $q \times q$ matrix and let $\lambda_1(y), \ldots, \lambda_q(y)$ denote the eigenvalues. Order them such that $\lambda_1(y) \equiv 1$. Then there is a $\tilde{y}$ and a sequence $\{y_n\} \subset [y_\ell, y_u]$ such that as $y_n \to \tilde{y}$, at least one eigenvalue (other than the one which is always unity) approaches the unit circle. In particular, suppose that the ordering is such that $|\lambda_2(y_n)| \to 1$ and that (choosing a subsequence if necessary) the $\lambda_i(y_n)$ converge to some $\tilde{\lambda}_i$ as $n \to \infty$, for $i = 1,\ldots,q$. The $\{\tilde{\lambda}_i\}$ must be the eigenvalues of $A(\tilde{y})$. But then $A'(y)$ is not the transition matrix of an ergodic process, a contradiction.

Q.E.D.

<u>Definition</u>. Let $\Sigma(y)$ denote the span of the eigenvectors and generalized eigenvectors of $A(y)$, except for the eigenvector which corresponds to the eigenvalue unity.

<u>Theorem 3</u>. <u>Assume the situation of Theorem</u> 1, <u>but let</u> $A(\cdot)$ <u>be continuously differentiable on</u> $[y_\ell, y_u]$ <u>(at the endpoints, take the left- or right-hand derivatives, as appropriate); then so is</u> $P(\cdot)$, <u>and</u> $P_y(y)$ <u>is the unique solution in</u> $\Sigma(y)$ <u>to the equation</u>

$$(4.1) \qquad P_y(y) = A(y)P_y(y) + A_y(y)P(y).$$

<u>Furthermore, the derivative</u> $P_{n,y}(y)$ <u>given by</u>

$$(4.2) \qquad P_{n+1,y}(y) = A(y)P_{n,y}(y) + A_y(y)P_n(y)$$

<u>converges geometrically to</u> $P_y(y)$, <u>uniformly in</u> $y \in [y_\ell, y_u]$ <u>and in the initial condition</u> $P_0(y)$, <u>if we set</u> $P_{0,y}(y) = 0$.

<u>If</u> $A(\cdot)$ <u>has continuous second derivatives on</u> $[y_\ell, y_u]$, <u>then so do</u> $P(\cdot)$ <u>and</u> $P_n(\cdot)$, <u>and</u> $P_{yy}(y)$ <u>is the unique solution in</u> $\Sigma(y)$ <u>to</u>

$$(4.3) \qquad P_{yy}(y) = A(y)P_{yy}(y) + 2A_y(y)P_y(y) + A_{yy}(y)P(y).$$

<u>Also</u>, $P_{n,yy}(y)$ <u>converges geometrically to</u> $P_{yy}(y)$, <u>uniformly in</u> $y \in [y_\ell, y_u]$ <u>and in the initial conditions, if</u> $P_{0,y}(y) = P_{0,yy}(y) = 0$.

<u>Proof</u>. Fix y. Since $(I-A(y))V = 0$ for $V \in \Sigma(y)$ implies that $V = 0$, in order for (4.1) to have a unique solution in $\Sigma(y)$ it is necessary and sufficient that $A_y(y)P(y) \perp \mathcal{N}(I-A'(y))$, where $\mathcal{N}$ denotes the null space of the matrix. $\mathcal{N}(I-A'(y))$ is the set of vectors Q such that $A'(y)Q = Q$. Since there is a unique eigenvalue of value unity and since the row <u>sums</u> of $A'(y)$ are all unity, the components of Q must all have the same value. Thus, the necessary and sufficient condition reduces to $A_y(y)P(y) \perp$ constant vectors. For any constant vector $C = (c,c,\ldots)'$, $C'A(y) = C'$. Thus, $C'A_y(y) = 0$ and hence $A_y(y)D \perp$ constant vectors for <u>any</u> vector D. Consequently (4.1) has a unique solution $\overline{P}_y(y)$ in $\Sigma(y)$.

Next, we show that $\overline{P}_y(y)$ is the desired derivative. Write (for $y \in (y_\ell, y_u)$, otherwise $\delta > 0$ or $\delta < 0$, as appropriate)

$$A(y+\delta)P(y+\delta) - A(y)P(y) = P(y+\delta) - P(y).$$

Thus,

$$\text{(4.4)} \qquad \frac{[A(y+\delta)-A(y)]}{\delta} P(y+\delta) = (I-A(y))\frac{[P(y+\delta)-P(y)]}{\delta}.$$

The left-hand side of (4.4) is uniformly bounded and is in $\Sigma(y)$ for each $\delta > 0$ (since $(I-A(y))V \in \Sigma(y)$ for any V) and it converges to $A_y(y)P(y)$ as $\delta \to 0$. When considered as an operator from $\Sigma(y)$ to $\Sigma(y)$, $[I-A(y)]$ has a bounded inverse. Thus, as $\delta \to 0$, $[P(y+\delta)-P(y)]/\delta$ converges to $P_y(y)$, which must equal $\overline{P}_y(y)$, by the uniqueness proved above.

We now turn to the convergence (4.2). By Theorem 1, $P_n(y)$ converges geometrically to $P(y)$, uniformly in $[y_\ell, y_u]$ and in $P_0(y)$. Also, since we use $P_{0,y}(y) = 0$,

$$P_{n+1,y}(y) = \sum_{j=0}^{n} A^{n-i}(y) A_y(y) P_i(y) .$$

But $A_y(y)P_i(y)$ is a bounded sequence in $\Sigma(y)$, and as $i \to \infty$ it converges geometrically and uniformly to $A_y(y)P(y)$. Also $A(y)$ is a contraction when acting in $\Sigma(y)$, uniformly in $y \in [y_\ell, y_u]$. These facts imply the desired convergence of $P_{n,y}(y)$. The limit must be a solution to (4.1).

The assertions concerning P_{yy} are proved in the same way and we omit the details. Q.E.D.

V. TIGHTNESS OF $\{U_n^\varepsilon$, SMALL ε, LARGE $n\}$

By "ε small" and "n large" we mean that there are $\varepsilon_0 > 0$, $N_\varepsilon < \infty$, such that the assertion holds for $\varepsilon \le \varepsilon_0$, $n \ge N_\varepsilon$. The actual value of ε_0 will be unimportant. Basic to the proof of weak convergence of $\{\tilde{U}^\varepsilon(\cdot)\}$ is the tightness of $\{U_n^\varepsilon$, small ε, large $n\}$.

<u>Theorem 4</u>. <u>For each small</u> $\varepsilon > 0$, <u>there is an</u> $N_\varepsilon < \infty$ <u>such that the doubly indexed sequence</u> $\{U_n^\varepsilon$, ε <u>small,</u> $n \ge N_\varepsilon\}$ <u>is tight, where</u> $U_n^\varepsilon = (y_n^\varepsilon - \bar{y})/\sqrt{\varepsilon}$.

<u>Proof</u>. Define $V(y) = (y-\bar{y})^2$. We have

$$(5.1a)\quad E_n^\varepsilon(y_{n+1}^\varepsilon - y_n^\varepsilon) = \mu\varepsilon[\alpha_\varepsilon(y_n^\varepsilon)y_n^\varepsilon(1-\nu_1 I\{X_n^{\varepsilon,1}=N_1\}) + \beta_\varepsilon(y_n^\varepsilon)(1-y_n^\varepsilon)(1-\nu_2 I\{X_n^{\varepsilon,2}=N_2\}],$$

$$(5.1b)\quad E_n^\varepsilon(y_{n+1}^\varepsilon - y_n^\varepsilon)^2 = \varepsilon^2\mu[\alpha_\varepsilon^2(y_n^\varepsilon)y_n^\varepsilon(1-\nu_1 I\{X_n^{\varepsilon,1}=N_1\}) + \beta_\varepsilon^2(y_n^\varepsilon)(1-y_n^\varepsilon)(1-\nu_2 I\{X_n^{\varepsilon,2}=N_2\})].$$

For small ε,

$$E_n^\varepsilon(y_n^\varepsilon-\bar{y})[\alpha_\varepsilon(y_n^\varepsilon)J_{1,n}^\varepsilon + \beta_\varepsilon(y_n^\varepsilon)J_{2,n}^\varepsilon] \le E_n^\varepsilon(y_n^\varepsilon-\bar{y})[\alpha(y_n^\varepsilon)J_{1,n}^\varepsilon + \beta(y_n^\varepsilon)J_{2,n}^\varepsilon],$$

since $0 \le \alpha_\varepsilon(y) \le \alpha(y)$ and $\alpha_\varepsilon(y) \ne \alpha(y)$ only if $y_n^\varepsilon - \bar{y} \ge 0$ (for small ε), and conversely for the β_ε term. Using the above inequality, (5.1a) and $|y_{n+1}^\varepsilon - y_n^\varepsilon| = O(\varepsilon)$,

$$(5.2)\quad E_n^\varepsilon V(y_{n+1}^\varepsilon) - V(y_n^\varepsilon) \le 2\mu\varepsilon(y_n^\varepsilon-\bar{y})[\alpha(y_n^\varepsilon)y_n^\varepsilon(1-\nu_1 I\{X_n^{\varepsilon,1}=N_1\})$$

$$+ \beta(y_n^\varepsilon)(1-y_n^\varepsilon)(1-\nu_2 I\{X_n^{\varepsilon,2}=N_2\})] + O(\varepsilon^2).$$

Define $V_1^\varepsilon(n)$ by

$$(5.3)\quad V_1^\varepsilon(n) = 2\mu\varepsilon(y_n^\varepsilon-\bar{y})\alpha(y_n^\varepsilon)y_n^\varepsilon\nu_1 \sum_{j=n}^{\infty}[P^1(N_1|y_n^\varepsilon) - P^1(X_n^\varepsilon,j-n,N_1|y_n^\varepsilon)]$$

$$+ 2\mu\varepsilon(y_n^\varepsilon-\bar{y})\beta(y_n^\varepsilon)(1-y_n^\varepsilon)\nu_2 \sum_{j=n}^{\infty}[P^2(N_2|y_n^\varepsilon) - P^2(X_n^\varepsilon,j-n,N_2|y_n^\varepsilon)].$$

Note that $P^i(X_n^\varepsilon,0,N_i|y_n^\varepsilon) = I\{X_n^{\varepsilon,i}=N_i\}$. By Theorem 2, the sums converge absolutely (the summands go to zero at a geometric rate) uniformly in n, y_n^ε, X_n^ε. Thus $|V_1^\varepsilon(\cdot)| = O(\varepsilon)$, uniformly in all the variables.

Next, evaluate

$$E_n^\varepsilon V_1^\varepsilon(n+1)-V_1^\varepsilon(n) = -2\mu\varepsilon(y_n^\varepsilon-\bar{y})\alpha(y_n^\varepsilon)y_n^\varepsilon\nu_1[P^1(N_1|y_n^\varepsilon) - I\{X_n^{\varepsilon,1}=N_1\}]$$

$$- 2\mu\varepsilon(y_n^\varepsilon-\bar{y})\beta(y_n^\varepsilon)(1-y_n^\varepsilon)\nu_2[P^2(N_2|y_n^\varepsilon) - I\{X_n^{\varepsilon,2}=N_2\}]$$

$$(5.4)\quad + \sum_{j=n+1}^{\infty} 2\mu\varepsilon\nu_1 \{E_n^\varepsilon(y_{n+1}^\varepsilon-\bar{y})\alpha(y_{n+1}^\varepsilon)y_{n+1}^\varepsilon[P^1(N_1|y_{n+1}^\varepsilon) - P^1(X_{n+1}^\varepsilon,j-n-1,N_1|y_{n+1}^\varepsilon)]$$

$$- (y_n^\varepsilon-\bar{y})\alpha(y_n^\varepsilon)y_n[P^1(N_1|y_n^\varepsilon) - P^1(X_n^\varepsilon,j-n,N_1|y_n^\varepsilon)]\}$$

+ a similar sum for route 2.

We next show that the sums in (5.4) = $O(\varepsilon^2)$ uniformly in all the variables n, y_n^ε, X_n^ε. For simplicity we work only with the first sum (route 1). By $|y_{n+1}^\varepsilon-y_n^\varepsilon| = O(\varepsilon)$, the smoothness of $\alpha(\cdot)$ and $\beta(\cdot)$ and Theorem 2, the sum changes by $O(\varepsilon^2)$ if $(y_{n+1}^\varepsilon-\bar{y})\alpha(y_{n+1}^\varepsilon)y_{n+1}^\varepsilon$ is replaced by $(y_n^\varepsilon-\bar{y})\alpha(y_n^\varepsilon)y_n^\varepsilon$. Upon making the substitution and using the Markov property of $\{X_j(y),\ j\geq n\}$ with the value $y = y_n^\varepsilon$ and "initial" condition

$$X_n(y_n^\varepsilon) = X_n^\varepsilon,$$

$$E_n^\varepsilon P^1(X_{n+1}^\varepsilon, j-n-1, N_1 | y_n^\varepsilon) = P^1(X_n^\varepsilon, j-n, N_1 | y_n^\varepsilon),$$

we can rewrite the sum as

$$(5.5) \quad O(\varepsilon) \sum_{j=n+1}^{\infty} E_n^\varepsilon \{ [P^1(N_1 | y_{n+1}^\varepsilon) - P^1(N_1 | y_n^\varepsilon)]$$

$$- [P^1(X_{n+1}^\varepsilon, j-n-1, N_1 | y_{n+1}^\varepsilon) - P^1(X_{n+1}^\varepsilon, j-n-1, N_1 | y_n^\varepsilon)] \} + O(\varepsilon^2).$$

Write $\delta y_n^\varepsilon = y_{n+1}^\varepsilon - y_n^\varepsilon$, and use the differentiability (Theorem 3) of the P^i and the law of the mean to write (5.5) in the form

$$O(\varepsilon)\delta y_n^\varepsilon \sum_{j=n+1}^{\infty} E_n^\varepsilon \int_0^1 [P_y^1(N_1 | y_n^\varepsilon + s\delta y_n^\varepsilon) - P_y^1(X_{n+1}^\varepsilon, j-n-1, N_1 | y_n^\varepsilon + s\delta y_n)] ds + O(\varepsilon^2).$$

By Theorem 3, the sequence of absolute values of the integrands converges to zero geometrically as $j \to \infty$, uniformly in s, n, δy_n^ε, and X_{n+1}^ε. This, together with $|\delta y_n^\varepsilon| = O(\varepsilon)$, yield that (5.5) is $O(\varepsilon^2)$. The same result holds for the sum in (5.4) corresponding to route 2.

Define $V^\varepsilon(n) = V(y_n^\varepsilon) + V_1^\varepsilon(n)$. By (5.2) and (5.4) and the fact that the sums in (5.4) are $O(\varepsilon^2)$,

$$E_n^\varepsilon V^\varepsilon(n+1) - V^\varepsilon(n) \le O(\varepsilon^2) + 2\mu\varepsilon(y_n^\varepsilon - \tilde{y})[\alpha(y_n^\varepsilon)y_n^\varepsilon(1-\nu_1 P^1(N_1 | y_n^\varepsilon))$$

$$+ \beta(y_n^\varepsilon)(1-y_n^\varepsilon)(1-\nu_2 P^2(N_2 | y_n^\varepsilon))].$$

Owing to the definition of $\alpha(\cdot)$ and $\beta(\cdot)$ and the fact that $y_n^\varepsilon \in [y_\ell, y_u]$, the

bracketed term has its unique zero at $y_n^\varepsilon = \bar{y}$ and it is positive (negative, resp.) for $y_n^\varepsilon < \bar{y}$ ($y_n^\varepsilon > \bar{y}$, resp.). Thus, there is a $\gamma > 0$ such that

$$(5.6) \qquad E_n^\varepsilon V^\varepsilon(n+1) - V^\varepsilon(n) \leq O(\varepsilon^2) - \varepsilon\gamma V(y_n^\varepsilon).$$

By $|V_1^\varepsilon(n)| = O(\varepsilon)$ uniformly in n, $E_n^\varepsilon V^\varepsilon(n+1) - V^\varepsilon(n) \leq O(\varepsilon^2) - \varepsilon\gamma V^\varepsilon(n)$, and hence

$$(5.7) \qquad EV^\varepsilon(n) \leq (\exp - \varepsilon\gamma n) EV^\varepsilon(0) + O(\varepsilon).$$

Again, since $|V_1^\varepsilon(n)| = O(\varepsilon)$, uniformly in n, (5.7) holds for $V(y_n^\varepsilon)$ replacing $V^\varepsilon(n)$, from which the existence of the $\{N_\varepsilon\}$ and the asserted tightness follows. In particular, let $0 < K_0$ be arbitrary and let N_ε be the smallest integer n such that $(\exp - \varepsilon n\gamma) \leq K_0\varepsilon$. Q.E.D.

VI. WEAK CONVERGENCE OF $\{\tilde{U}^{\varepsilon}(\cdot)\}$

Definition. Recall the definition of N_ε given at the end of the proof of Theorem 4. For any sequence of integers $n_\varepsilon > N_\varepsilon$, define $\rho_\varepsilon = n_\varepsilon - N_\varepsilon$. Define $\tilde{y}^\varepsilon_n = y^\varepsilon_{n_\varepsilon + n}$ and similarly define the "shifted" sequences $\tilde{U}^\varepsilon_n$, $\tilde{X}^\varepsilon_n$ and $\tilde{J}^\varepsilon_{in}$. Then

$$\tilde{U}^\varepsilon_{n+1} = \tilde{U}^\varepsilon_n + \sqrt{\varepsilon}[\alpha_\varepsilon(\tilde{y}^\varepsilon_n)\tilde{J}^\varepsilon_{1n} + \beta_\varepsilon(\tilde{y}^\varepsilon_n)\tilde{J}^\varepsilon_{2n}]. \tag{6.1}$$

By Theorem 4, $\{\tilde{U}^\varepsilon_n, \varepsilon \text{ small}\}$ is tight. For each integer N, define $\tilde{U}^{\varepsilon,N}_n$, $\tilde{y}^{\varepsilon,N}_n$, $\tilde{J}^{\varepsilon,N}_{in}$ via

$$\tilde{U}^{\varepsilon,N}_{n+1} = \tilde{U}^{\varepsilon,N}_n + \sqrt{\varepsilon}[\alpha(\tilde{y}^{\varepsilon,N}_n)\tilde{J}^{\varepsilon,N}_{1n} + \beta(\tilde{y}^{\varepsilon,N}_n)\tilde{J}^{\varepsilon,N}_{2n}]b_N(\tilde{U}^{\varepsilon,N}_n), \tag{6.2}$$

where $b_N(\cdot)$ is defined above (2.1) and we set $\tilde{U}^{\varepsilon,N}_0 = \tilde{U}^\varepsilon_0$ if $|\tilde{U}^\varepsilon_0| \le N$ and equal to zero otherwise. Also $\tilde{U}^{\varepsilon,N}_n = (\tilde{y}^{\varepsilon,N}_n - \bar{y})/\sqrt{\varepsilon}$ defines $\tilde{y}^{\varepsilon,N}_n$. $\tilde{J}^{\varepsilon,N}_{in}$ is simply the indicator function of the set {route i is tried first and call accepted} for the system $\{\tilde{X}^{\varepsilon,N}_n, \tilde{y}^{\varepsilon,N}_n\}$, where the choice probabilities $\{\tilde{y}^{\varepsilon,N}_n\}$ are used to select the routes and $\{\tilde{X}^{\varepsilon,N}_n\}$ is the corresponding route occupancy process. We suppose that $\tilde{X}^{\varepsilon,N}_0 = X^\varepsilon_{n_\varepsilon}$. Let $\tilde{E}^{\varepsilon,N}_n$ denote expectation conditional on $\tilde{y}^{\varepsilon,N}_n$ and $\tilde{X}^{\varepsilon,N}_n$. Since $|\tilde{y}^{\varepsilon,N}_n - \bar{y}| \le \sqrt{\varepsilon}(N+1)$, for small ε it is irrelevant whether we use α_ε, β_ε or α, β in (6.2), and we use α, β for simplicity. By Theorem 1, if we show that (for each N) $\{\tilde{U}^{\varepsilon,N}(\cdot)\}$ is tight and that all weak limits satisfy (6.3) until first escape from S_N, then $\{\tilde{U}^\varepsilon(\cdot)\}$ is tight and all weak limits satisfy (6.3).

We now define some auxiliary processes which are used in the averaging method employed in the proof. Let $\bar{P}$ denote the measure defined by the stationary process $\{X_j(\bar{y}), \infty > j > -\infty\}$, with corresponding expectation operator $\bar{E}$. For each n, it is necessary to introduce the process $\{X_j(\bar{y}), j \ge n\}$, but with "initial" condition $X_n(\bar{y}) = \tilde{X}^{\varepsilon,N}_n$. (I.e., after time n, the route choice probability is $\bar{y}$.) The opera-

tor $\bar{E}_n^{\varepsilon,N}$ denotes the expectation of functions of this process $\{X_j(\bar{y}), j\geq n\}$ conditional on the "initial" condition $X_n(\bar{y}) = \tilde{X}_n^{\varepsilon,N}$. Let $J_{ij}(\bar{y})$ denote the indicator function I{call arrives at j+1, is assigned to and accepted by route i}, when the route choice variable is $\bar{y}$ and the route occupancy process is $\{X_j(\bar{y})\}$. Whether we intend the ergodic process or the process $\{X_j(\bar{y}), j\geq n\}$ starting at time n with $X_n(\bar{y}) = \tilde{X}_n^{\varepsilon,N}$ will be made obvious by use of either $\bar{E}$ or $\bar{E}_n^{\varepsilon,N}$. Define

$$\delta u_j(\bar{y}) = [\alpha(\bar{y})J_{1j}(\bar{y}) + \beta(\bar{y})J_{2j}(\bar{y})].$$

Under $\bar{P}$, the right side has zero expectation.

<u>Theorem 5</u>. <u>For any sequence</u> $n_\varepsilon \geq N_\varepsilon$, $\{\tilde{U}^\varepsilon(\cdot)\}$ <u>is tight in</u> $D[0,\infty)$. <u>All weakly convergent subsequences converge to a Gauss-Markov diffusion satisfying</u> (6.3). <u>If</u> $\varepsilon Q_\varepsilon \to \infty$ <u>as</u> $\varepsilon \to \infty$, <u>then the limiting diffusion</u> $u(\cdot)$ <u>is stationary in that</u> $u(0)$ <u>has the stationary distribution</u>. (<u>In all cases</u> $u(0)$ <u>is independent of</u> $B(\cdot)$.)

(6.3) $\quad du = Gudt + \sigma dB, \quad B(\cdot) =$ <u>standard Brownian motion</u>,

(6.4) $$G = \hat{F}_y(\bar{y}) = \frac{\partial}{\partial y}\mu y(1-y)[\nu_2 P^2(N_2|y) - \nu_1 P^1(N_1|y)]\Big|_{y=\bar{y}},$$

(6.5) $$\sigma^2 = \bar{E}(\delta u_0(\bar{y}))^2 + 2\sum_{n=1}^{\infty}\bar{E}\,\delta u_0(\bar{y})\delta u_n(\bar{y}).$$

<u>Proof</u>. <u>Part 1</u>. <u>Until Part 4, all superscripts</u> N <u>will be omitted</u>. <u>Thus we write</u> $\{\tilde{E}_n^\varepsilon, \bar{E}_n^\varepsilon, \tilde{X}_n^\varepsilon, \tilde{y}_n^\varepsilon, \tilde{X}_n^{\varepsilon,i}, \ldots\}$ <u>for</u> $\{\tilde{E}_n^{\varepsilon,N}, \bar{E}_n^{\varepsilon,N}, \tilde{X}_n^{\varepsilon,N}, \tilde{y}_n^{\varepsilon,N}, \tilde{X}_n^{\varepsilon,N,i}, \ldots\}$. <u>We actually work with the</u> N-<u>truncated process</u> in Parts 1 to 3.

By (5.1),

$$\tilde{E}_n^{\varepsilon}(\tilde{U}_{n+1}^{\varepsilon}-\tilde{U}_n^{\varepsilon}) = \sqrt{\varepsilon}\mu\tilde{y}_n^{\varepsilon}(1-\tilde{y}_n^{\varepsilon})[\nu_2 I\{\tilde{X}_n^{\varepsilon,2}=N_2\}-\nu_1 I\{\tilde{X}_n^{\varepsilon,1}=N_1\}]b_N(\tilde{U}_n^{\varepsilon}). \tag{6.6}$$

Let $f(\cdot,\cdot) \in \mathcal{D} = \mathcal{L}_0^{2,3}$, the space of bounded (x,t) functions with compact support whose mixed partial derivatives up to order 2 in t and 3 in x are continuous. To apply Theorem 1 to $\{\tilde{U}^{\varepsilon}(\cdot)\}$, we will get an $f^{\varepsilon}(\cdot)$ of the form

$$f^{\varepsilon}(n\varepsilon) = f(\tilde{U}_n^{\varepsilon},n\varepsilon) + f_0^{\varepsilon}(n\varepsilon) + f_1^{\varepsilon}(n\varepsilon) + f_2^{\varepsilon}(n\varepsilon)$$

where the $f_i^{\varepsilon}(n\varepsilon)$ will be defined in the sequel. For each N, all $o(\cdot)$ or $O(\cdot)$ are uniform in all variables except their argument. We have

$$\tilde{E}_n^{\varepsilon}f(\tilde{U}_{n+1}^{\varepsilon},n\varepsilon+\varepsilon)-f(\tilde{U}_n^{\varepsilon},n\varepsilon) = \tilde{E}_n^{\varepsilon}[f(\tilde{U}_{n+1}^{\varepsilon},n\varepsilon)-f(\tilde{U}_n^{\varepsilon},n\varepsilon)] + f_t(\tilde{U}_n^{\varepsilon},n\varepsilon)\varepsilon + o(\varepsilon),$$

$$\tilde{E}_n^{\varepsilon}[f(\tilde{U}_{n+1}^{\varepsilon},n\varepsilon)-f(\tilde{U}_n^{\varepsilon},n\varepsilon)] = \tilde{E}_n^{\varepsilon}f_u(\tilde{U}_n^{\varepsilon},n\varepsilon)(\tilde{U}_{n+1}^{\varepsilon}-\tilde{U}_n^{\varepsilon}) + \frac{1}{2}\tilde{E}_n^{\varepsilon}f_{uu}(\tilde{U}_n^{\varepsilon},n\varepsilon)(\tilde{U}_{n+1}^{\varepsilon}-\tilde{U}_n^{\varepsilon})^2 + o(\varepsilon)$$

$$\begin{aligned} &= \sqrt{\varepsilon}\mu f_u(\tilde{U}_n^{\varepsilon},n\varepsilon)\tilde{y}_n^{\varepsilon}(1-\tilde{y}_n^{\varepsilon})b_N(\tilde{U}_n^{\varepsilon})[\nu_2 I\{\tilde{X}_n^{\varepsilon,2}=N_2\}-\nu_1 I\{\tilde{X}_n^{\varepsilon,1}=N_1\}] \\ &\quad + \frac{f_{uu}(\tilde{U}_n^{\varepsilon},n\varepsilon)}{2}\tilde{E}_n^{\varepsilon}(\tilde{U}_{n+1}^{\varepsilon}-\tilde{U}_n^{\varepsilon})^2 + o(\varepsilon). \end{aligned} \tag{6.7}$$

By the differentiability result of Theorem 3, we can rewrite the term before the $o(\varepsilon)$ as follows:*

$$\begin{aligned} &\varepsilon b_N^2(\tilde{U}_n^{\varepsilon})\frac{f_{uu}(\tilde{U}_n^{\varepsilon},n\varepsilon)}{2}\tilde{E}_n^{\varepsilon}[\alpha(\tilde{y}_n^{\varepsilon})\tilde{J}_{1n}^{\varepsilon}+\beta(\tilde{y}_n^{\varepsilon})\tilde{J}_{2n}^{\varepsilon}]^2 \\ &\qquad = \varepsilon b_N^2(\tilde{U}_n^{\varepsilon})\frac{f_{uu}(\tilde{U}_n^{\varepsilon},n\varepsilon)}{2}\bar{E}_n^{\varepsilon}[\alpha(\bar{y})J_{1n}(\bar{y})+\beta(\bar{y})J_{2n}(\bar{y})]^2 + o(\varepsilon). \end{aligned}$$

---

*The terms $\bar{E}_n^{\varepsilon}J_{1n}(\bar{y})$ and $\tilde{E}_n^{\varepsilon}J_{1n}^{\varepsilon}(\tilde{y}_n^{\varepsilon})$ differ only in that in the first case $\bar{y}$ is used as the choice variable to get the successor state to $\tilde{X}_n^{\varepsilon}$, and $\tilde{y}_n^{\varepsilon}$ is used in the second case.

Part 2. We will "average out" the terms in (6.7) one by one. Define $f_1^\varepsilon(n\varepsilon)$ (analogous to the definition of $V_1(n)$ in the last section)

$$(6.8)\quad f_1^\varepsilon(n\varepsilon) = \sqrt{\varepsilon}\mu b_N(\tilde{U}_n^\varepsilon)\tilde{y}_n^\varepsilon(1-\tilde{y}_n^\varepsilon)f_u(\tilde{U}_n^\varepsilon,n\varepsilon)\sum_{j=n}^{\infty}[\nu_2(P^2(\tilde{X}_n^\varepsilon,j-n,N_2|\tilde{y}_n^\varepsilon) -$$

$$- P^2(N_2|\tilde{y}_n^\varepsilon)) - \nu_1(P^1(\tilde{X}_n^\varepsilon,j-n,N_1|\tilde{y}_n^\varepsilon) - P^1(N_1|\tilde{y}_n^\varepsilon))].$$

Proceeding analogously to the method of Theorem 4 for $V_1^\varepsilon(n)$, we evaluate (writing $P^i(\tilde{X}_n^\varepsilon,j-n,N_i|\tilde{y}_n^\varepsilon)$ in the more convenient form $\tilde{E}_n^\varepsilon P^i(\tilde{X}_{n+1}^\varepsilon,j-n-1,N_i|\tilde{y}_n^\varepsilon)$ in T_3 below, for $j > n$; see above (5.5))

$$(6.9)\quad \tilde{E}_n^\varepsilon f_1^\varepsilon(n\varepsilon+\varepsilon)-f_1^\varepsilon(n\varepsilon) = T_1 + T_2 + T_3,$$

$$T_1 = -\sqrt{\varepsilon}\mu\tilde{y}_n^\varepsilon(1-\tilde{y}_n^\varepsilon)b_N(\tilde{U}_n^\varepsilon)f_u(\tilde{U}_n^\varepsilon,n\varepsilon)[(\nu_2 I\{\tilde{X}_n^{\varepsilon,2}=N_2\}-\nu_1 I\{\tilde{X}_n^{\varepsilon,1}=N_1\})$$

$$- (\nu_2 P^2(N_2|\tilde{y}_n^\varepsilon)-\nu_1 P^1(N_1|\tilde{y}_n^\varepsilon))]$$

$$T_2 = \sqrt{\varepsilon}\mu\tilde{E}_n^\varepsilon[f_u(\tilde{U}_{n+1}^\varepsilon,n\varepsilon+\varepsilon)b_N(\tilde{U}_{n+1}^\varepsilon)\tilde{y}_{n+1}^\varepsilon(1-\tilde{y}_{n+1}^\varepsilon)]\cdot$$

$$\cdot\sum_{j=n+1}^{\infty}[\nu_2(P^2(\tilde{X}_{n+1}^\varepsilon,j-n-1,N_2|\tilde{y}_{n+1}^\varepsilon)-P^2(N_2|\tilde{y}_{n+1}^\varepsilon))$$

$$-\nu_1(P^1(\tilde{X}_{n+1}^\varepsilon,j-n-1,N_1|\tilde{y}_{n+1}^\varepsilon)-P^1(N_1|\tilde{y}_{n+1}^\varepsilon))],$$

$$T_3 = -\sqrt{\varepsilon}\mu[f_u(\tilde{U}_n^\varepsilon,n\varepsilon)\tilde{y}_n^\varepsilon(1-\tilde{y}_n^\varepsilon)b_N(\tilde{U}_n^\varepsilon)]\cdot$$

$$\cdot\sum_{j=n+1}^{\infty}\tilde{E}_n^\varepsilon[\nu_2(P^2(\tilde{X}_{n+1}^\varepsilon,j-n-1,N_2|\tilde{y}_n^\varepsilon)-P^2(N_2|\tilde{y}_n^\varepsilon))$$

$$-\nu_1(P^1(\tilde{X}_{n+1}^\varepsilon,j-n-1,N_1|\tilde{y}_n^\varepsilon)-P^1(N_1|\tilde{y}_n^\varepsilon))].$$

Using the differentiability result of Theorem 3 and the fact that $\nu_2 P^2(N_2|\bar{y}) = \nu_1 P^1(N_1|\bar{y})$, we get that T_1 equals the negative of the first term on the right side of (6.7) plus

$$(6.10)\quad \varepsilon\mu b_N(\tilde{U}_n^\varepsilon)\frac{\partial}{\partial y}\{y(1-y)[\nu_2 P^2(N_2|y)-\nu_1 P^1(N_1|y)]\tilde{U}_n^\varepsilon\}\Big|_{y=\bar{y}} + o(\varepsilon).$$

In T_2, by replacing $\tilde{y}_{n+1}^\varepsilon$ by $\tilde{y}_n^\varepsilon$ and $b_N(\tilde{U}_{n+1}^\varepsilon)f_u(\tilde{U}_{n+1}^\varepsilon, n\varepsilon+\varepsilon)$ by

$$b_N(\tilde{U}_n^\varepsilon)f_u(\tilde{U}_n^\varepsilon, n\varepsilon) + (b_N(\tilde{U}_n^\varepsilon)f_u(\tilde{U}_n^\varepsilon, n\varepsilon))_u(\tilde{U}_{n+1}^\varepsilon - \tilde{U}_n^\varepsilon),$$

<u>we only alter the term by</u> $o(\varepsilon)$. Let us make these replacements in T_2 and denote the resulting term by T_2^0. Now, split T_2^0 into two parts (T_{21}^2, T_{22}^2), the first (second, resp.) being T_2 but with $b_N(\tilde{U}_n^\varepsilon)f_u(\tilde{U}_n^\varepsilon, n\varepsilon)$ $((b_N(\tilde{U}_n^\varepsilon)f_u(\tilde{U}_n^\varepsilon, n\varepsilon))_u(\tilde{U}_{n+1}^\varepsilon - \tilde{U}_n^\varepsilon)$, resp.) replacing $b_N(\tilde{U}_{n+1}^\varepsilon)f_u(\tilde{U}_{n+1}^\varepsilon, n\varepsilon+\varepsilon)$. By the differentiability results of Theorem 3 and the fact that $|\tilde{y}_{n+1}^\varepsilon - \tilde{y}_n^\varepsilon| = O(\varepsilon)$ and an argument like that below (5.5), it can be shown that $T_{21}^0 + T_3 = o(\varepsilon)$. Thus

$$(6.11a)\quad T_2 + T_3 = o(\varepsilon) + \sqrt{\varepsilon}\mu\tilde{y}_n^\varepsilon(1-\tilde{y}_n^\varepsilon)(b_N(\tilde{U}_n^\varepsilon)f_u(\tilde{U}_n^\varepsilon, n\varepsilon))_u$$

$$\cdot\, \tilde{E}_n^\varepsilon(\tilde{U}_{n+1}^\varepsilon - \tilde{U}_n^\varepsilon)\sum_{j=n+1}^{\infty}[\nu_2(P^2(\tilde{X}_{n+1}^\varepsilon, j-n-1, N_2|\tilde{y}_n^\varepsilon) - P^2(N_2|\tilde{y}_n^\varepsilon))$$

$$-\ \nu_1(P^1(\tilde{X}_{n+1}^\varepsilon, j-n-1, N_1|\tilde{y}_n^\varepsilon) - P^1(N_1|\tilde{y}_n^\varepsilon))].$$

We now simplify (6.11a) by a series of replacements, each one altering the term by $o(\varepsilon)$. First replace all the $\tilde{y}_n^\varepsilon$ by $\bar{y}$. By Theorem 3 and $|\tilde{U}_{n+1}^\varepsilon - \tilde{U}_n^\varepsilon| = O(\sqrt{\varepsilon})$ and a differentiability argument such as used below (5.5), this only alters the

term by $o(\varepsilon)$. Since $\nu_2 P^2(N_2|\bar{y}) - \nu_1 P^1(N_1|\bar{y}) = 0$, we delete this part of the resulting summand. We now have

$$(6.11b) \qquad T_2 + T_3 = o(\varepsilon) + \sqrt{\varepsilon}\,[b_N(\tilde{U}_n^\varepsilon) f_u(\tilde{U}_n^\varepsilon, n\varepsilon)]_u \tilde{E}_n^\varepsilon (\tilde{U}_{n+1}^\varepsilon - \tilde{U}_n^\varepsilon) \sum_{j=n+1}^{\infty} q_j^\varepsilon,$$

where for $j \geq n+1$,

$$q_j^\varepsilon = [\nu_2 P^2(\tilde{X}_{n+1}^\varepsilon, j-n-1, N_2|\bar{y}) - \nu_1 P^1(\tilde{X}_{n+1}^\varepsilon, j-n-1, N_1|\bar{y})]\mu\bar{y}(1-\bar{y}) = \bar{E}_{n+1}^\varepsilon \delta u_j(\bar{y}).$$

Finally, by the differentiability result of Theorem 3, (6.11b) equals

$$(6.12) \quad T_2 + T_3 = o(\varepsilon) + \varepsilon b_N(\tilde{U}_n^\varepsilon)[b_N(\tilde{U}_n^\varepsilon) f_u(\tilde{U}_n^\varepsilon, n\varepsilon)]_u \sum_{j=n+1}^{\infty} \bar{E}_n^\varepsilon \delta u_n(\bar{y}) \delta u_j(\bar{y}).$$

The difference between (6.11b) and (6.12) is simply due to whether $\tilde{y}_n^\varepsilon$ or $\bar{y}$ is used to get $\tilde{X}_{n+1}^\varepsilon$ and $\tilde{U}_{n+1}^\varepsilon$ from $\tilde{X}_n^\varepsilon$ and $\tilde{U}_n^\varepsilon$.

<u>Part 3</u>. Now, we "average out" the sum in (6.12). Define $f_2^\varepsilon(n\varepsilon)$ by

$$f_2^\varepsilon(n\varepsilon) = \varepsilon b_N(\tilde{U}_n^\varepsilon)[b_N(\tilde{U}_n^\varepsilon) f_u(\tilde{U}_n^\varepsilon, n\varepsilon)]_u \sum_{j=n}^{\infty} \sum_{k=j+1}^{\infty} [\bar{E}_n^\varepsilon \delta u_j(\bar{y}) \delta u_k(\bar{y}) - \bar{E}\delta u_j(\bar{y}) \delta u_k(\bar{y})].$$

By the (uniform) geometric convergence result of Theorem 2, the sum converges absolutely and $|f_2^\varepsilon(n\varepsilon)| = O(\varepsilon)$. By a straightforward calculation using the stationarity of $\{\delta u_n(\bar{y})\}$ under $\bar{P}$, we can show that

$$\tilde{E}_n^\varepsilon f_2^\varepsilon(n\varepsilon+\varepsilon) - f_2^\varepsilon(n\varepsilon) = -\,(6.12) + o(\varepsilon)$$

$$+\ \varepsilon b_N(\tilde{U}_n^\varepsilon)[b_N(\tilde{U}_n^\varepsilon) f_u(\tilde{U}_n^\varepsilon, n\varepsilon)]_u \sum_{j=1}^{\infty} \bar{E}(\delta u_0(\bar{y}) \delta u_j(\bar{y})).$$

Finally, we treat the term before the $o(\varepsilon)$ of (6.7) - in the form in which it is written below (6.7). Define $f_0^\varepsilon(n\varepsilon)$ by

$$f_0^\varepsilon(n\varepsilon) = \varepsilon \frac{f_{uu}(\tilde{U}_n^\varepsilon, n\varepsilon)}{2} b_N^2(\tilde{U}_n^\varepsilon) \sum_{j=n}^{\infty} [\overline{E}_n^\varepsilon(\delta u_j(\bar{y}))^2 - \overline{E}(\delta u_j(\bar{y}))^2].$$

By a procedure similar to that used for $f_1^\varepsilon(n\varepsilon)$, it can readily be shown that

$$\tilde{E}_n^\varepsilon f_0^\varepsilon(n\varepsilon+\varepsilon) - f_0^\varepsilon(n\varepsilon) = o(\varepsilon) + \varepsilon \frac{f_{uu}(\tilde{U}_n^\varepsilon, n\varepsilon)}{2} b_N^2(\tilde{U}_n^\varepsilon)\overline{E}(\delta u_0(\bar{y}))^2$$

$$- \varepsilon \frac{f_{uu}(\tilde{U}_n^\varepsilon, n\varepsilon)}{2} b_N^2(\tilde{U}_n^\varepsilon)\overline{E}_n^\varepsilon[\alpha(\bar{y})J_{1n}(\bar{y}) + \beta(\bar{y})J_{2n}(\bar{y})]^2.$$

Summarizing the previous calculations

$$\tilde{E}_n^\varepsilon f^\varepsilon(n\varepsilon+\varepsilon) - f^\varepsilon(n\varepsilon) = o(\varepsilon) + \varepsilon f_t(\tilde{U}_n^\varepsilon, n\varepsilon) + \varepsilon f_u(\tilde{U}_n^\varepsilon, n\varepsilon) G\tilde{U}_n^\varepsilon b_N(\tilde{U}_n^\varepsilon)$$

$$(6.14) \qquad + \varepsilon f_u(\tilde{U}_n^\varepsilon, n\varepsilon) b_{N,u}(\tilde{U}_n^\varepsilon) b_N(\tilde{U}_n^\varepsilon) \sum_{j=1}^{\infty} \overline{E}\, \delta u_0(\bar{y}) \delta u_j(\bar{y})$$

$$+ \varepsilon \frac{f_{uu}(\tilde{U}_n^\varepsilon, n\varepsilon)}{2} b_N^2(\tilde{U}_n^\varepsilon) [\overline{E}(\delta u_0(\bar{y}))^2 + 2\sum_{j=1}^{\infty} \overline{E}\, \delta u_0(\bar{y}) \delta u_j(\bar{y})].$$

<u>Part 4</u>. <u>Conclusion</u>. <u>Reintroduce the superscript</u> N. <u>Fix</u> N. All the $f_i^{\varepsilon,N}$ are bounded and of order $O(\sqrt{\varepsilon})$ and $\{\tilde{U}_0^{\varepsilon,N}\} = \{\tilde{U}^{\varepsilon,N}(0)\}$ is tight. Also $\tilde{E}_n^{\varepsilon,N} f^{\varepsilon,N}(n\varepsilon+\varepsilon) - f^{\varepsilon,N}(n\varepsilon) = O(\varepsilon)$. Thus, by [7], Theorem 2, the bounded sequence $\{\tilde{U}^{\varepsilon,N}(\cdot)\}$ is tight in $D[0,\infty)$. Let ε index a weakly convergent subsequence with limit $U^N(\cdot)$. Since A is defined to be the infinitesimal operator of (6.3), by (6.14) and Theorem 1, we see that $U^N(\cdot)$ solves the martingale problem corresponding to an infinitesimal operator A^N whose coefficients equal those of A in S_N. Thus, by Theorem 1, $\{\tilde{U}^\varepsilon(\cdot)\}$ converges weakly to a solution $u(\cdot)$ of (6.3). The independence of B(.) and u(0) is a consequence of the fact that $u(\cdot)$ is the unique solution to the martingale problem. The stationarity assertion is not hard to prove, but we omit the details. Q.E.D.

VII. ASYMPTOTIC THEORY OF AN ADAPTIVE QUANTIZER: INTRODUCTION

In recent years there has been a great deal of effort concerning the efficient quantization of signals in telecommunications systems, e.g. of voice signals in telephone transmission systems. Let $z(\cdot)$ denote the actual signal process and Δ a sampling interval. In the problem of interest, the signal is sampled at moments $\{n\Delta, n=0,1,\ldots\}$, then the samples $\{z(n\Delta)\}$ are quantized, and it is only the quantized samples which are transmitted. Let $0 = \xi_0 < \xi_1 < \ldots < \xi_{L-1} < \xi_L = \infty$, $0 = \eta_1 < \eta_2 \ldots < \eta_L$ where ξ_i, η_{i+1}, $i = 0, \ldots, L-1$, are real numbers. Let the quantization function $Q(\cdot)$ be defined as follows: there is a $y > 0$ such that for $z(n\Delta) > 0$, $Q(z(n\Delta)) = y\eta_i$ if $z(n\Delta) \in [y\xi_{i-1}, y\xi_i)$, and set $Q(-z) = -Q(z)$. The parameter y is a scaling parameter. As the signal power increases (decreases), y should increase (decrease) for efficient reconstruction of the signal from the sequence of quantizations.

The problem of choosing appropriate values of y when the signal powers can vary by an order of magnitude or more has led to the study of adaptive quantizers. We give only a brief description in order to formulate the problem. For more detail and discussion of the engineering considerations, the reader is referred to the references [4], [5]. Let ε denote a "rate of adjustment" parameter for the scale parameter y and let y_n^ε denote the value of the adapted scale parameter at the nth sampling instant. Set $\beta \in (0,1]$ and let $0 < M_1^\varepsilon < M_2^\varepsilon < \ldots < M_L^\varepsilon < \infty$ with $M_1^\varepsilon < 1$, $M_L^\varepsilon > 1$. We study an adaptive quantizer which is a truncated form of the (typical in such an application) adaptive system

$$(7.1) \qquad y_{n+1}^\varepsilon = (y_n^\varepsilon)^\beta B_n^\varepsilon, \quad \text{where } B_n^\varepsilon = M_i^\varepsilon \text{ if } |z(n\Delta)| \in [y_n^\varepsilon \xi_{i-1}, y_n^\varepsilon \xi_i).$$

Goodman and Gersho [4] did a thorough analysis of (7.1) for the case $\beta = 1$ and $\{z(n\Delta)\}$ independent and identically distributed. With $\beta < 1$, the system

has some desirable robustness properties and this case, together with simulations, is discussed by Mitra [5] and others. The last reference is concerned more with reconstruction of the process $z(\cdot)$ from $\{Q(z(n\Delta))\}$ and does not give an asymptotic analysis.

Generally, with non-i.i.d. $\{z(n\Delta)\}$, it is hard to get concrete information on $\{y_n^\varepsilon\}$ for large n. If the signal power varies over time or if (as is realistic for moderate values of Δ) $\{z(n\Delta)\}$ is not i.i.d., then techniques such as used in [4] fail, but for small rates of adjustment (ε) an asymptotic analysis can still shed light on the process behavior. At the present time, it seems that little more can be done for the general case. Here, we scale the problem so that an asymptotic analysis is possible. For mathematical as well as practical purposes, it is useful to confine y_n^ε to some finite positive interval $[y_\ell, y_u]$. Now, we define the truncated form of (7.1) which will be studied. Let $\alpha > 0$, $0 < \alpha\varepsilon < 1$ and let $\{\ell_i\}$ be real numbers such that $\ell_1 < \ell_2 < \dots < \ell_L$ and $\ell_1 < 0$, $\ell_L > 0$. Then we use

$$y_{n+1}^\varepsilon = (y_n^\varepsilon)^{1-\varepsilon\alpha} B_n^\varepsilon\Big|_{y_\ell}^{y_u}, \tag{7.2}$$

where $|$ denotes truncation and

$$B_n^\varepsilon = (1+\varepsilon\ell_i) \quad \text{if} \quad |z(n\Delta)| \in [y_n^\varepsilon \xi_{i-1}, y_n^\varepsilon \xi_i).$$

The asymptotic results can be used to get information on the effects of the $\{\ell_i\}$, Δ, structure of $z(\cdot)$ and α on the performance for small ε. For notational convenience below, let $y_\ell < 1$ and $y_u > 1$. Rewrite (7.2) in the form (7.3), where $y^{1-\varepsilon\alpha} = y[1-\varepsilon\alpha \log y]+O(\varepsilon^2)$ and $(1+\varepsilon b_n^\varepsilon) \equiv B_N^\varepsilon$ are used, and F and b_n^ε have the obvious definitions.

$$(7.3) \quad y_{n+1}^{\varepsilon} = [y_n^{\varepsilon}(1+\varepsilon b_n^{\varepsilon})-\varepsilon\alpha y_n^{\varepsilon} \log y_n^{\varepsilon}+O(\varepsilon^2)]\Big|_{y_\ell}^{y_u} \equiv [y_n^{\varepsilon}+\varepsilon F(y_n^{\varepsilon},z(n\Delta))+O(\varepsilon^2)]\Big|_{y_\ell}^{y_u}.$$

In [4], the process $\{\log y_n^{\varepsilon}\}$ rather than $\{y_n^{\varepsilon}\}$ is dealt with.

We proceed in very much the same way that we did for the automata problem. The main difference arises from the unboundedness of $\{z(n\Delta)\}$, under assumption (7.6). By definition,

$$b_n^{\varepsilon} = \sum_{i=1}^{L} \ell_i I\{|z(n\Delta)| \in [y_n^{\varepsilon}\xi_{i-1}, y_n^{\varepsilon}\xi_i)\}.$$

There are continuous functions $\ell_i^{\varepsilon}(\cdot)$ such that (7.4) and the properties below it hold.

$$(7.4) \qquad y_{n+1}^{\varepsilon} = y_n^{\varepsilon}(1+\varepsilon\beta_n^{\varepsilon}(y_n^{\varepsilon})) - \varepsilon\alpha y_n^{\varepsilon} \log y_n^{\varepsilon} + O(\varepsilon^2)$$

$$\equiv y_n^{\varepsilon} + \varepsilon F_{\varepsilon}(y_n^{\varepsilon}, z(n\Delta)) + O(\varepsilon^2),$$

where

$$(7.5) \qquad \beta_n^{\varepsilon}(y) = \sum_{i=1}^{L} \ell_i^{\varepsilon}(y) I\{|z(n\Delta)| \in [y\xi_{i-1}, y\xi_i)\ .$$

Also, $\ell_i^{\varepsilon}(\cdot)$ can be chosen such that $\ell_i^{\varepsilon}(\cdot) = \ell_i$ out of an $O(\varepsilon)$ neighborhood of y_ℓ (resp. y_u) if $\ell_i < 0$ (resp. $\ell_i > 0$), and $0 \geq \ell_i^{\varepsilon}(y) \geq \ell_i$ for $\ell_i < 0$ and $0 \leq \ell_i^{\varepsilon}(y) \leq \ell_i$ for $\ell_i > 0$.

Some assumptions. For specificity, $z(\cdot)$ is assumed to be a stationary Gaussian process with a rational spectral density. Thus there are an asymptotically stable matrix M, a matrix C, a row vector D, and a process $v(\cdot)$ such that

$$(7.6)\qquad dv = Mvdt + Cdw$$

$$z = Dv, \quad w(\cdot) = \text{vector-valued standard Brownian motion.}$$

This assumption is not essential - only certain smoothness properties of the multivariate density are used, together with the exponential rate of decrease of the effects of the initial conditions.

Define $\hat{F}_\varepsilon(y) = EF_\varepsilon(y,z(n\Delta))$ and $\hat{F}(y) = EF(y,z(n\Delta))$. Let $\sigma_0^2 = \text{var } z(t)$. We have (the subscript y denotes the derivative)

$$(7.7)\quad \frac{d}{dy}\left(\frac{\hat{F}(y)}{y}\right) = \frac{2}{\sqrt{2\pi}\sigma_0}\sum_{i=1}^{L} \ell_i[\xi_i \exp - \frac{\xi_i^2 y^2}{2\sigma_0^2} - \xi_{i-1} \exp - \frac{\xi_{i-1}^2 y^2}{2\sigma_0^2}] - \alpha/y$$

$$= \frac{2}{\sqrt{2\pi}\sigma_0}\sum_{i=1}^{L-1} (\ell_i - \ell_{i+1})\xi_i \exp -\xi_i^2 y^2/2\sigma_0^2 - \alpha/y.$$

We can see from the terms in (7.7) that $\hat{F}(y)/y$ is the sum of two strictly convex functions, the first being bounded and having a negative slope, and the second going to ∞ as $y \to 0$ and to $-\infty$ as $y \to \infty$. Thus there is a unique $\bar{y} \in (0,\infty)$ such that $\hat{F}_y(\bar{y}) = 0$. Also $\hat{F}(y) > 0$ for $0 < y < \bar{y}$ and $\hat{F}(y) < 0$ for $y > y$ and $\hat{F}_y(\bar{y}) \neq 0$. We assume that $y \in (y_\ell, y_u)$. For small ε, the assertions in the last sentence hold with $\hat{F}_\varepsilon$ replacing $\hat{F}$. Define $U_n^\varepsilon = (y_n^\varepsilon - \bar{y})/\sqrt{\varepsilon}$ and let E_n denote expectation conditioned on $\{v(j\Delta), j<n\}$.

VIII. TIGHTNESS OF $\{U_n^\varepsilon$, SMALL ε, LARGE $n\}$

The proof is similar to that of Theorem 4 in Section V and we only set it up and indicate how to deal with the fact that $\{z(n\Delta)\}$ is unbounded.

Theorem 6. Under the conditions in Section VII, the conclusions of Theorem 4 hold.

Proof. Define $V(y) = (y-\bar{y})^2$. There is a $\gamma > 0$ such that $(y-\bar{y})\hat{F}(y) \le -\gamma V(y)$, all $\varepsilon > 0$ and $y \in [y_\ell, y_u]$. We have

$$(y_{n+1}^\varepsilon - y_n^\varepsilon)^2 = O(\varepsilon^2), \qquad y_{n+1}^\varepsilon = y_n^\varepsilon + \varepsilon\hat{F}_\varepsilon(y_n^\varepsilon) + \varepsilon[F_\varepsilon(y_n^\varepsilon, z(n\Delta)) - \hat{F}_\varepsilon(y_n^\varepsilon)] + O(\varepsilon^2),$$

$$\hat{F}_\varepsilon(y) = y\sum_{i=1}^{L} \ell_i^\varepsilon(y) P\{y\xi_{i-1} \le |z(n\Delta)| < y\xi_i\} - \alpha(y \log y),$$

(8.1)

$$E_n^\varepsilon(y_{n+1}^\varepsilon - y_n^\varepsilon) = \varepsilon\hat{F}_\varepsilon(y_n^\varepsilon) + \varepsilon\, y_n^\varepsilon \sum_{i=1}^{L} \ell_i^\varepsilon(y_n^\varepsilon)[P\{y\xi_{i-1} \le |z(n\Delta)| < y\xi_i | v(n\Delta - \Delta)\}$$

$$- P\{y\,\xi_{i-1} \le |z(n\Delta)| < y\,\xi_i\}]\Big|_{y=y_n^\varepsilon} + O(\varepsilon^2).$$

As done in connection with (5.2) (where α_ε, β_ε were replaced by α,β), we get an upper bound for the second moment by replacing $\ell_i(y_n^\varepsilon)$ by ℓ_i (hence $\hat{F}_\varepsilon$ by $\hat{F}$). Thus

(8.2) $$E_n^\varepsilon V(y_{n+1}^\varepsilon) - V(y_n^\varepsilon) \le O(\varepsilon^2) + 2\varepsilon(y_n^\varepsilon - \bar{y})\hat{F}(y_n^\varepsilon)$$

$$+ 2(y_n^\varepsilon - \bar{y})[\text{sum in (8.1) with } \ell_i^\varepsilon(\cdot) \text{ replaced by } \ell_i].$$

Next, define $V_1^\varepsilon(n)$ by $V_1^\varepsilon(n) = V_1^\varepsilon(n,y_n^\varepsilon)$, where

$$(8.3)\qquad V_1^\varepsilon(n,y) = 2\varepsilon\,(y-\bar{y}) \sum_{j=n}^{\infty} [\sum_{i=1}^{L} y\ell_i\,[P\{y\xi_{i-1}\leq|z(j\Delta)|<y\xi_i\,|\,v(n\Delta-\Delta)\}$$

$$- P\{y\xi_{i-1}\leq|z(n\Delta)|<y\xi_i\}].$$

$|V_1^\varepsilon(n)|$ can be estimated by use of the following fact. There are $K_0 < \infty$ and $a > 0$ such that $|e^{Mt}| \leq K_0 e^{-at}$. There is an $a_1 > 0$ and a $K_1 < \infty$ such that for $\tau_2 > \tau_1 > 0$ and on the set $\{v(t): |v(t)|e^{-a\tau_1/2} \leq 1\}$,

$$(8.4)\qquad |P\{v(t+\tau_i)\in B_i,\ i=1,2|v(t)\} - P\{v(t+\tau_i)\in B_i,\ i=1,2\}| \leq K_1 e^{-a_1\tau_1}.$$

for all B_1, B_2.

In order to use (8.4) (in this application we set B_2 = range space of $v(t)$), write the sum in (8.3) as

$$(8.5)\qquad \sum_{j=n}^{H} + \sum_{j=H+1}^{\infty},$$

where $H = \min\{m: e^{-(m-n)\Delta a/2}|v(n\Delta-\Delta)|\leq 1\} = O(1+\max(0, \log|v(n\Delta-\Delta)|))$. Then the first sum in (8.5) is $O(1+\max(0, \log|v(n\Delta-\Delta)|))$, and the second is $O(1)$ by (8.4) and the summability of $\sum_{j\geq 0}\exp -a_1 j\Delta$. Thus $|V_1^\varepsilon(n)| = O(\varepsilon)[1+\max(0, \log|v(n\Delta-\Delta)|)] \leq O(\varepsilon)(1+|v(n\Delta-\Delta)|)$. From this point on, the proof is exactly the same as that for Theorem 4. Q.E.D.

IX. THE LIMIT THEOREM

We continue to use the tilde $\sim$ terminology of Section VI, and define $\tilde{U}_n^\varepsilon$, $\tilde{y}_n^\varepsilon$, $\tilde{E}_n^\varepsilon$, etc., as there. Also, set $\tilde{z}(n\Delta) = z(n_\varepsilon\Delta+n\Delta)$ and $\tilde{v}(n\Delta) = v(n_\varepsilon\Delta+n\Delta)$. The idea now is still to prove weak convergence of $\tilde{U}^\varepsilon(\cdot)$. We use $\tilde{E}_n^\varepsilon$ for expectation conditional on $\{v(j\Delta),\ j<n+n_\varepsilon\}$. We have ((9.1b) defines $\tilde{y}_n^{\varepsilon,N}$ by $\tilde{U}_n^{\varepsilon,N} = (\tilde{y}_n^{\varepsilon,N}-\bar{y})/\sqrt{\varepsilon}$)

$$(9.1a)\qquad \tilde{U}_{n+1}^\varepsilon = \tilde{U}_n^\varepsilon + \sqrt{\varepsilon}\hat{F}_\varepsilon(\tilde{y}_n^\varepsilon) + \sqrt{\varepsilon}(F_\varepsilon(\tilde{y}_n^\varepsilon,\tilde{z}(n\Delta))-\hat{F}_\varepsilon(\tilde{y}_n^\varepsilon)) + O(\varepsilon^{3/2}),$$

$$(9.1b)\qquad \tilde{U}_{n+1}^{\varepsilon,N} = \tilde{U}_n^{\varepsilon,N} + \sqrt{\varepsilon}[\hat{F}_\varepsilon(\tilde{y}_n^{\varepsilon,N}) + (F_\varepsilon(\tilde{y}_n^{\varepsilon,N},\tilde{z}(n\Delta)) - \hat{F}_\varepsilon(\tilde{y}_n^{\varepsilon,N})) + O(\varepsilon^{3/2})]b_N(\tilde{U}_n^{\varepsilon,N}).$$

Theorem 7. Under the conditions of Section VII, the conclusions of Theorem 5 hold, but where $G = \hat{F}_y(\bar{y})$ and (stationary process $z(\cdot)$ used)

$$\sigma^2 = EF^2(\bar{y},z(0)) + 2\sum_{n=1}^{\infty} EF(\bar{y},z(n\Delta))F(\bar{y},z(0)).$$

Remark. If M, C or D were time-varying, then an extension of the technique is possible, provided that the time variation per step is $O(\varepsilon)$. The limit diffusion yields information on the dependence of the performance on the parameters α, $\{\ell_i\}$, Δ, $\{\xi_i\}$, as well as an estimate of the asymptotic variance and correlation function for small ε.

Proof. Except for the unboundedness of the noise $\{z(n\Delta)\}$, the proof would be essentially the same as that of Theorem 5, and only an outline will be given.

Owing to the truncation $|\tilde{U}_n^{\varepsilon,N}| \le N+1$, the F_ε, $\hat{F}_\varepsilon$ in (9.1b) can be replaced by F and $\hat{F}$, respectively, without changing the values, for small ε. Let us make the replacement. Fix $f(\cdot,\cdot)\in \hat{\mathcal{C}}_0^{2,3}$. Drop the superscript N on all variables for notational convenience, as done in Theorem 5. Then, by a Taylor expansion,

$$(9.2)\quad \tilde{E}_n^{\varepsilon} f(\tilde{U}_{n+1}^{\varepsilon}, n\varepsilon+\varepsilon) - f(\tilde{U}_n^{\varepsilon}, n\varepsilon) = o(\varepsilon) + \varepsilon f_t(\tilde{U}_n^{\varepsilon}, n\varepsilon) + \varepsilon f_u(\tilde{U}_n^{\varepsilon}, n\varepsilon)\hat{F}_y(\bar{y})\tilde{U}_n^{\varepsilon} b_N(\tilde{U}_n^{\varepsilon})$$

$$+ \sqrt{\varepsilon} f_u(\tilde{U}_n^{\varepsilon}, n\varepsilon)\tilde{E}_n^{\varepsilon}[F(\bar{y}+\sqrt{\varepsilon}\tilde{U}_n^{\varepsilon}, \tilde{z}(n\Delta)) - \hat{F}(\bar{y}+\sqrt{\varepsilon}\tilde{U}_n^{\varepsilon})]b_N(\tilde{U}_n^{\varepsilon})$$

$$+ \frac{\varepsilon}{2} f_{uu}(\tilde{U}_n^{\varepsilon}, n\varepsilon)\tilde{E}_n^{\varepsilon}[F(\bar{y}+\sqrt{\varepsilon}\tilde{U}_n^{\varepsilon}, \tilde{z}(n\Delta)) - \hat{F}(\bar{y}+\sqrt{\varepsilon}\tilde{U}_n^{\varepsilon})]^2 b_N^2(\tilde{U}_n^{\varepsilon})).$$

Since the second derivative of $\tilde{E}_n^{\varepsilon}F(y, \tilde{z}(n\Delta))$ with respect to y is bounded by constant$[1+|\tilde{v}(n\Delta-\Delta)|]$, the next-to-last term of (9.2) can be written as

$$(9.3)\quad \sqrt{\varepsilon} f_u(\tilde{U}_n^{\varepsilon}, n\varepsilon)\tilde{E}_n^{\varepsilon}[F(\bar{y}, \tilde{z}(n\Delta)) - \hat{F}(\bar{y})]b_N(\tilde{U}_n^{\varepsilon})$$

$$+ \varepsilon f_u(\tilde{U}_n^{\varepsilon}, n\varepsilon)\frac{\partial}{\partial y}\tilde{E}_n^{\varepsilon}[F(y, \tilde{z}(n\Delta)) - \hat{F}(y)]\Big|_{y=\bar{y}} \tilde{U}_n^{\varepsilon} b_N(\tilde{U}_n^{\varepsilon}) + o(\varepsilon)[1+|\tilde{v}(n\Delta-\Delta)|].$$

The last term of (9.2) can be written as (recall that $\hat{F}(\bar{y}) = 0$)

$$(9.4)\quad \frac{\varepsilon}{2} f_{uu}(\tilde{U}_n^{\varepsilon}, n\varepsilon)\tilde{E}_n^{\varepsilon}[F(\bar{y}, \tilde{z}(n\Delta)) - \hat{F}(\bar{y})]^2 b_N^2(\tilde{U}_n^{\varepsilon}) + o(\varepsilon).$$

Now, we use the method of Theorem 5 in order to average out the terms of (9.2). We use $f^{\varepsilon}(n\varepsilon) = f(\tilde{U}_n^{\varepsilon}, n\varepsilon) + \sum_{i=3}^{6} f_i^{\varepsilon}(n\varepsilon)$. Define $f_3^{\varepsilon}(n\varepsilon)$ by (to average out the second term of (9.3))

$$f_3^{\varepsilon}(n\varepsilon) = \varepsilon f_u(\tilde{U}_n^{\varepsilon}, n\varepsilon) b_N(\tilde{U}_n^{\varepsilon})\tilde{U}_n^{\varepsilon} \sum_{j=n}^{\infty} \frac{\partial}{\partial y} \tilde{E}_n^{\varepsilon}[F(y, \tilde{z}(j\Delta)) - \hat{F}(y)]\Big|_{y=\bar{y}}.$$

By an argument similar to that used below (8.5), together with the derivative bound stated above (9.3), it can be shown that $\tilde{E}_n^{\varepsilon} f_3^{\varepsilon}(n\varepsilon+\varepsilon) - f_3^{\varepsilon}(n\varepsilon) = -$ (second term of (9.3)) $+ o(\varepsilon)[1 + |\tilde{v}(n\Delta-\Delta)|^2]$ and that $|f_3^{\varepsilon}(n\varepsilon)| \leq O(\varepsilon)[1 + |\tilde{v}(n\Delta-\Delta)|]$. Next, introduce $f_4^{\varepsilon}(n\varepsilon)$ (to average out (9.4)):

$$f_4^\varepsilon(n\varepsilon) = \frac{\varepsilon}{2} f_{uu}(\tilde{U}_n^\varepsilon, n\varepsilon) b_N^2(\tilde{U}_n^\varepsilon) \sum_{j=n}^{\infty} [\tilde{E}_n^\varepsilon F^2(\bar{y}, \tilde{z}(j\Delta)) - EF^2(\bar{y}, \tilde{z}(j\Delta))].$$

Then, as for f_3^ε, we have $|f_4^\varepsilon(n\varepsilon)| \le O(\varepsilon)[1 + |\tilde{v}(n\Delta-\Delta)|]$. Using this, it is not hard to show via a small amount of manipulation that

$$\tilde{E}_n^\varepsilon f_4^\varepsilon(n\varepsilon+\varepsilon) - f_4^\varepsilon(n\varepsilon) = -\frac{\varepsilon}{2} f_{uu}(\tilde{U}_n^\varepsilon, n\varepsilon) b_N(\tilde{U}_n^\varepsilon) [\tilde{E}_n^\varepsilon F^2(\bar{y}, \tilde{z}(n\Delta)) - EF^2(\bar{y}, \tilde{z}(n\Delta))]$$

$$+ o(\varepsilon)[1 + |\tilde{v}(n\Delta-\Delta)|].$$

Next, introduce $f_5^\varepsilon(n\varepsilon)$ in order to average out the first term of (9.3):

$$f_5^\varepsilon(n\varepsilon) = \sqrt{\varepsilon}\, f_u(\tilde{U}_n^\varepsilon, n\varepsilon) b_N(\tilde{U}_n^\varepsilon) \sum_{j=n}^{\infty} \tilde{E}_n^\varepsilon F(\bar{y}, \tilde{z}(j\Delta)).$$

Then, again, $|f_5^\varepsilon(n\varepsilon)| = o(\sqrt{\varepsilon})(1 + |\tilde{v}(n\Delta-\Delta)|)$ and we can write

(9.5a) $$\tilde{E}_n^\varepsilon f_5^\varepsilon(n\varepsilon+\varepsilon) - f_5^\varepsilon(n\varepsilon) = -\text{ (first term of (9.3)}$$

$$+ \varepsilon \tilde{E}_n^\varepsilon [f_u(\tilde{U}_{n+1}^\varepsilon, n\varepsilon) b_N(\tilde{U}_{n+1}^\varepsilon) - f_u(\tilde{U}_n^\varepsilon, n\varepsilon) b_N(\tilde{U}_n^\varepsilon)] \sum_{j=n+1}^{\infty} \tilde{E}_{n+1}^\varepsilon F(\bar{y}, \tilde{z}(j\Delta)).$$

With a small amount of manipulation, we can show that the last term of (9.5a) equals

(9.5b) $$\varepsilon b_N(\tilde{U}_n^\varepsilon)[f_u(\tilde{U}_n^\varepsilon, n\varepsilon) b_N(\tilde{U}_n^\varepsilon)]_u \sum_{j=n+1}^{\infty} \tilde{E}_n^\varepsilon F(\bar{y}, \tilde{z}(j\Delta)) F(\bar{y}, \tilde{z}(n\Delta)) + o(\varepsilon)[1 + |\tilde{v}(n\Delta-\Delta)|].$$

Finally, $f_6^\varepsilon(n\varepsilon)$ is introduced in order to average out the sum term in (9.5b) in the same way that $f_2^\varepsilon(n\varepsilon)$ was used to average out (6.12) in Theorem 5. Define

$$(9.6)\quad f_6^\varepsilon(n\varepsilon) = \varepsilon[f_u(\tilde{U}_n^\varepsilon,n\varepsilon)b_N(\tilde{U}_n^\varepsilon)]_u b_N(\tilde{U}_n^\varepsilon)\cdot$$

$$\cdot\sum_{j=n}^{\infty}\sum_{k=j+1}^{\infty}[\tilde{E}_n^\varepsilon F(\bar{y},\tilde{z}(k\Delta))F(\bar{y},\tilde{z}(j\Delta))-EF(\bar{y},\tilde{z}(k\Delta))F(\bar{y},\tilde{z}(j\Delta))].$$

By (8.4), $f_6^\varepsilon(n\varepsilon)$ is well defined and is $O(\varepsilon)[1+|\tilde{v}(n\Delta-\Delta)|^2]$, as will now be proved.

Define H as below (8.5) and let $\tilde{E}_n^\varepsilon B_{jk}^\varepsilon$ denote the (j,k)th <u>summand</u> in (9.6) and write the sum in (9.6) as

$$\sum_{j=n}^{H}\sum_{k=j+1}^{\infty}\tilde{E}_n^\varepsilon B_{jk}^\varepsilon+\sum_{j=H+1}^{\infty}\sum_{k=j+1}^{\infty}\tilde{E}_n^\varepsilon B_{jk}^\varepsilon \equiv I + II.$$

By the argument connected with (8.5), the inner sum in I is bounded by

$$(9.7)\quad \sum_{k=j+1}^{\infty}|\tilde{E}_n^\varepsilon B_{jk}^\varepsilon| \le \sum_{k=j+1}^{\infty}\tilde{E}_n^\varepsilon|\tilde{E}_j^\varepsilon B_{jk}| = O(1+|\tilde{v}(n\Delta-\Delta)|).$$

Thus, by the bound on (H-n), $I \le O(1+|\tilde{v}(n\Delta-\Delta)|^2)$. To treat II, we note the following: there is a $K_2<\infty$ such that for $H<j<k$, $|\tilde{E}_n^\varepsilon B_{jk}^\varepsilon| \le K_2\exp -a_1(j-n)\Delta$. Also, for $k>j$,

$$\tilde{E}_j^\varepsilon F(\bar{y},\tilde{z}(k\Delta)) \le K_2[\exp -a_1(k-j)\Delta+I\{|(\exp -a(k-j)\Delta)\tilde{v}(j\Delta-\Delta)| \ge 1\}].$$

With a little more work, these estimates yield the existence of a $K_3<\infty$ such that $|\tilde{E}_n^\varepsilon B_{jk}^\varepsilon| \le (1+O(|\tilde{v}(n\Delta-\Delta)|)K_3\exp -a_1\Delta[(j-n)+(k-j)]/2$, from which the fact that II = O(1) and the last sentence of the previous paragraph both follow.

It is straightforward to show that

$$\tilde{E}_n^{\epsilon} f_6^{\epsilon}(n\epsilon+\epsilon) - f_6^{\epsilon}(n\epsilon) = -(\text{sum term in } (9.5))$$

$$+ \epsilon b_N(\tilde{U}_n^{\epsilon})[f_u(\tilde{U}_n^{\epsilon}, n\epsilon) b_N(\tilde{U}_n^{\epsilon})]_u \sum_{n=1}^{\infty} EF(\bar{y}, \tilde{z}(n\Delta))F(\bar{y}, \tilde{z}(0)) + o(\epsilon)[1+|v(n\Delta-\Delta)|^2].$$

Summarizing, with $f^{\epsilon}(n\epsilon)$ defined by $f^{\epsilon}(n\epsilon) = f(\tilde{U}_n^{\epsilon}, n\epsilon) + \sum_{i=3}^{6} f_i^{\epsilon}(n\epsilon)$, we have

$$(9.8) \quad \tilde{E}_n^{\epsilon} f^{\epsilon}(n\epsilon+\epsilon) - f^{\epsilon}(n\epsilon) = o(\epsilon)[1+|\tilde{v}(n\Delta-\Delta)|^2) + \epsilon f_t(\tilde{U}_n^{\epsilon}, n\epsilon) + \epsilon f_u(\tilde{U}_n^{\epsilon}, n\epsilon)\hat{F}_y(\bar{y})\tilde{U}_n^{\epsilon} b_N(\tilde{U}_n^{\epsilon})$$

$$+ \epsilon b_N(\tilde{U}_n^{\epsilon})[f_u(\tilde{U}_n^{\epsilon}, n\epsilon) b_N(\tilde{U}_n^{\epsilon})]_u \sum_{n=1}^{\infty} EF(\bar{y}, z(n\Delta))F(\bar{y}, z(0))$$

$$+ \epsilon b_N(\tilde{U}_N^{\epsilon}) \frac{f_{uu}(\tilde{U}_n^{\epsilon}, n\epsilon)}{2} EF^2(\bar{y}, z(0)) + o(\epsilon)[1+|\tilde{v}(n\Delta-\Delta)|^2).$$

Now, if the $\{\tilde{U}^{\epsilon,N}(\cdot)\}$ (returning to the use of superscript N) were tight for each N, then (9.8) and Theorem 1 imply that any weakly convergent subsequence of $\{\tilde{U}^{\epsilon,N}(\cdot)\}$ converges to a diffusion with operator A^N, whose coefficients equal those of A in S_N and, hence, that the original $\{\tilde{U}^{\epsilon}(\cdot)\}$ converge weakly to the solution of (6.3) with the G and σ defined in Theorem 7.

But (dropping the superscript N again) $|\sum_{i=3}^{6} f_i^{\epsilon}(n\epsilon)| = O(\sqrt{\epsilon})[1+|\tilde{v}(n\Delta-\Delta)|^2]$ and $|\tilde{E}_n^{\epsilon} f^{\epsilon}(n\epsilon+\epsilon) - f^{\epsilon}(n\epsilon)| = O(\epsilon) + o(\epsilon)[1+|\tilde{v}(n\Delta-\Delta)|^2]$ and for any $T < \infty$, $K > 0$, the Gaussian property implies that

$$\lim_{\epsilon\to 0} P\{\sup_{n\le T/\epsilon} \epsilon|v(n\epsilon)|^2 \le K\} = 0.$$

Thus, tightness follows by Theorem 2 of [1] or [7], as it did for the case of Theorem 1. Q.E.D.

References

1. Kushner, H.J., Hai Huang (1979), "On the weak convergence of a sequence of general stochastic difference equations to a diffusion", submitted to SIAM J. on Applied Math.

2. Narendra, K.S., Wright, E.A., Mason, L.E. (1977), "Application of learning automata to telephone traffic routing and control", IEEE Trans. on Systems, Man and Cybernetics, SMC-7, 785-792.

3. Narendra, K.S., Thathachar, M.A.L., (1979), "On the behavior of a learning automaton in a changing environment with application to telephone traffic routing", preprint, Yale University, Dept. of Engineering.

4. Goodman, D.J., Gersho, A. (1974), "Theory of an adaptive quantizer", IEEE Trans. on Communications, COM-22, 1037-1045.

5. Mitra, D. (1979), "A generalized adaptive quantization system with a new reconstruction method for noisy transmission", IEEE Trans. on Communications, COM-27, 1681-1689.

6. Billingsley, P. (1968), Convergence of Probability Measures, John Wiley and Sons, New York.

7. Kushner, H.J. (1979), "A martingale method for the convergence of a sequence of processes to a jump-diffusion process", to appear Z. Wahrscheinlichkeitsteorie.

8. Strook, D.W., Varadhan, S.R.S. (1979), Multidimensional Diffusion Processes, Springer, Berlin.

9. Norman, M.F. (1974), "Markovian learning processes", SIAM Rev., 16, 143-162.

ANALYSIS OF NONLINEAR STOCHASTIC SYSTEMS WITH WIDE-BAND INPUTS

Harold J. Kushner*

Y. Bar-Ness**

Abstract

A general method for approximating systems with wide-band inputs by diffusion processes is discussed. The input noise itself can be processed nonlinearly by elements of the system. One particular case is examined in detail, that of a hard limiter just after the input. The results suggest that the limiter can actually improve performance if the noise intensity is small. The development illustrates how tricky, but potentially useful, nonlinearities can be handled when the system input is wide-band noise.

*Divisions of Applied Mathematics and Engineering, Lefschetz Center for Dynamical Systems, Brown University; work supported by AFOSR AF-76-3063, NSF Eng. 77-12946, ONR N00014-76-C-0279-P003.

**Division of Applied Mathematics, Lefschetz Center for Dynamical Systems, Brown University; on leave from Department of Electrical Engineering, University of Tel Aviv. Work supported by AFOSR AF-76-3063, ONR N00014-76-C-0279-P003.

I. Introduction

Consider a (linear or nonlinear) dynamical system with a wide-band noise input. It is often of considerable interest to approximate such systems by diffusion models so that, e.g., Markov process techniques can be used. In [1] - [4], [7], several powerful methods for doing this have been developed. Roughly, the input noise process is parametrized by ε and as $\varepsilon \to 0$, the bandwidth (BW) $\to \infty$, while the power per unit BW converges to a constant. The limit process is found via methods of weak convergence theory. The methods are particularly useful when the system noise (and/or signal) is processed nonlinearly; i.e., only nonlinear functions of the noise appear in the dynamics. The problem is often not what the so-called correction term might be, but what the entire form of the limit is, and this is not usually easy. In fact, when nonlinear functions of the noise appear, the notion of "correction term" loses much of its sense.

In this paper, the system of Figure 1 is dealt with.

$$
\begin{aligned}
&\dot{v}^{\varepsilon} = F(v^{\varepsilon}) + Dy^{\varepsilon}, \\
&y^{\varepsilon} = L_{\varepsilon} \operatorname{sign} u^{\varepsilon}, \qquad u^{\varepsilon} = s + n^{\varepsilon} - G(v^{\varepsilon}), \qquad v^{\varepsilon}(t) \in R^{r},
\end{aligned}
\tag{1.1}
$$

where $n^{\varepsilon}(\cdot)$ is a scalar-valued wide-band noise input process. Conditions on $F(\cdot)$, $n^{\varepsilon}(\cdot)$ and $G(\cdot)$ will be given below. The main result is that as $\varepsilon \to 0$ (BW $\to \infty$), the measures of $\{v^{\varepsilon}(\cdot)\}$ converge to those of $v(\cdot)$ where $v(\cdot)$ satisfies the Itô equation

$$
dv = F(v)dt + LD[(\frac{s-G(v)}{\sigma})(\sqrt{2/\pi})\ dt + \sqrt{2 \ln 2/a}\ dB],
\tag{1.2}
$$

where $B(\cdot)$ is a standard Brownian motion and L, σ, a will be defined below. Roughly, "a" is related to the correlation function of the $n^{\varepsilon}(\cdot)/(E|n^{\varepsilon}(t)|^2)^{1/2}$, and σ^2 is the intensity of the spectrum of $n^{\varepsilon}(\cdot)$ in any band [0,BW] for small ε.

If, in the system of Fig. 1, the saturator and gain L_ϵ were replaced by a gain K, then the limit would be (follows from the method of this paper)

$$(1.3) \qquad dv = F(v)dt + KD[(s-G(v))dt + \sqrt{2R}\, db],$$

where for the noise model used for (1.1), $R = \sigma^2/a$. There is no so-called "correction term". Owing to the form of the saturator function, the formal technique of Stratonovich is inapplicable.

The example is offered to illustrate what can be done with one particularly annoying but useful nonlinearity. The basic method is widely applicable. The scheme is unrelated to statistical linearization, which in fact is not concerned directly with approximating processes.

Before proceeding, compare (1.2) and (1.3) for the case when the feedback $-G(\cdot)$ is supposed to be stabilizing (i.e., when the system is designed to make the error $s(t)-G(v^\epsilon(t))$ small. In (1.2), the term in the dynamics which involves the error is proportional to $1/\sigma$, and in (1.3), the noise term is proportional to σ. Thus for small σ, we expect the limiter to enhance stability without increasing the noise effects, an important point to note. For large σ, the limiter does not seem to be helpful. A simulation comparison of the "pre-limit" with the limit for a somewhat different problem (a phase-locked loop with a saturator) suggests that the limit ($\epsilon \to 0$) results are often the "worst" case, in that (for example) the limit mean square error often increased to the limit value as $\epsilon \to 0$. (They also suggest that often the limit process is approached quite fast (as measured, say, by the mean square value of the input to the limiter) as $\epsilon \to 0$.) We do not know the extent of applicability of this rule - but it seems to hold frequently. When it does hold, the limit results can provide useful upper bounds, and system improvements suggested by the form of the limit might well be improvements

for the "pre-limit" case also. Unfortunately, it is not usually possible to get approximate diffusion processes where the BW is not large - so, even if it is not large, the results for large BW might be a useful guide to the qualitative behavior.

Reference [5] contains some applications to problems in communications theory of the same general idea. But owing to the unbounded nature of the noise and the form of the discontinuity and feedback, the problem here is harder and the analytical details different.

Section II gives specific assumptions. The main background theorem and some comments on weak convergence appear in Section III, and the convergence $\{v^\varepsilon(\cdot)\} \to v(\cdot)$ is proved in Section IV. A similar method would be used with other nonlinearities.

II. Model Assumptions

1. The noise model. Let $z(\cdot)$ denote a stationary Gaussian process with correlation function $\sigma^2 \exp -a|\tau|$, $a > 0$, and set $n^\varepsilon(t) = z^\varepsilon(t)/\varepsilon$, where $z^\varepsilon(t) = z(t/\varepsilon^2)$. As $\varepsilon \to 0$, the spectral density of $n^\varepsilon(\cdot)$ converges to $2\sigma^2/a$ on any finite interval. The scaling is a convenient and common way of getting a noise process $n^\varepsilon(\cdot)$ whose spectrum converges (as $\varepsilon \to 0$) to that of a white noise with a constant power/unit bandwidth. For other correlation functions the $\sqrt{2 \ln 2}$ in (1.2) is replaced by something slightly different. We use the noise form only to facilitate the evaluation of the coefficient of $dB(\cdot)$ in (1.2). The Gaussian assumption simplifies the proof that certain integrals converge - but is not essential.

2. The limiter gain L_ε. If $L_\varepsilon \equiv L$, a number not depending on ε, then as $\varepsilon \to 0$, the "increased wildness" of $n^\varepsilon(\cdot)$ essentially wipes out the saturator -

replacing it by an open circuit. Thus L_ε must increase as ε decreases. In any particular fixed practical system, one particular value of L_ε will be used. But as the bandwidth $\to \infty$, this value of L_ε will have to increase (see proof in Section IV) and $\varepsilon L_\varepsilon$ will have to converge to a non-zero number. So we use $L_\varepsilon = L/\varepsilon$.

3. _Other assumptions_. $G(\cdot)$, $F(\cdot)$ are continuously differentiable and the solution to (1.2) is unique in the weak sense. $s(\cdot)$ is right continuous and uniformly bounded on $[0,\infty)$. The method is most easy to use if the functions are smooth. _The analysis will be done with_ $g_\alpha(\cdot)$ _replacing_ $g(\cdot) = \text{sat}(\cdot)$, where the piecewise linear $g_\alpha(\cdot)$ is defined in Fig. 2. _We then get the result_ $\{v^\varepsilon(\cdot)\} \to v(\cdot)$ _as_ $\varepsilon \to 0$, _then_ $\alpha \to 0$.

III. Weak Convergence; A Convergence Theorem

Tightness. Let $D^r[0,\infty)$ denote the space of R^r-valued functions on $[0,\infty)$ which are right continuous and have left-hand limits. A certain topology called the Skorokhod topology ([6], section 14) is usually put on D^r. The process $v\ (\cdot)$ is considered to be a random variable with values in $D^r[0,\infty)$ and induces a measure P_ε on it. $\{P_\varepsilon\}$ or $\{v^\varepsilon(\cdot)\}$ is said to be _tight_ iff for each $\delta > 0$ there is a compact $K_\delta \in D^r[0,\infty)$ such that $P_\varepsilon(K_\delta) \geq 1-\delta$, all ε. $\{v^\varepsilon(\cdot)\}$ is said to _converge weakly_ to a process $v(\cdot)$ with paths in $D^r[0,\infty)$ and inducing measure P on it iff for each bounded real-valued function $g(\cdot)$ on $D^r[0,\infty)$, $\int q(\omega)dP_\varepsilon(\omega) \to \int q(\omega)dP(\omega)$ as $\varepsilon \to 0$. Thus weak convergence is a generalization of convergence in distribution. It is the appropriate form of convergence for our problem. The tightness condition for $\{v^\varepsilon(\cdot)\}$ will hold under our assumptions.

Truncated processes. The actual technical proofs of tightness and weak convergence are easier if the processes $\{v^{\varepsilon}(\cdot)\}$ are bounded. Define

$$\text{(3.1)}\qquad \dot v^{\varepsilon,N} = [F(v^{\varepsilon,N})+Dy^{\varepsilon,N}]b_N(v^{\varepsilon,N}), \qquad y^{\varepsilon,N} = L_{\varepsilon}g_{\alpha}(u^{\varepsilon,N}),$$

$$u^{\varepsilon,N} = s + n^{\varepsilon} - G(v^{\varepsilon,N}),$$

where $b_N(v) = 1$ for $v \in S_N = \{v: |v| \le N\}$, $b_N(v) = 0$ for $v \notin S_{N+1}$ and $b_N(v) \in [0,1]$ and has third derivatives that are bounded uniformly in v and N. If we can prove convergence for $\{v^{\varepsilon,N}(\cdot), \varepsilon\to 0\}$ for each N, then Theorem 1 says that we can prove it for (1.1). Thus, the truncation is purely technical and does not affect the result.

Definitions. Let A denote the infinitesimal operator of the diffusion (1.2). Let $\mathscr{F}_t^{\varepsilon,N}$ denote the σ-algebra induced by $\{v^{\varepsilon,N}(s), n^{\varepsilon}(s), s\le t\}$ and $E_t^{\varepsilon,N}$ the corresponding conditional expectation. Actually $\mathscr{F}_t^{\varepsilon,N}$ and $E_t^{\varepsilon,N}$ depend on α also. But we usually suppress the α affix. Let $\mathscr{M}$ be the class of measurable (ω,t) functions such that if $g(\cdot) \in \mathscr{M}$, then $E|g(t+\delta)-g(t)| \to 0$ as $\delta \downarrow 0$ and $\sup_t E|g(t)| < \infty$ and $g(t)$ depends only on $\{v^{\varepsilon,N}(s), n^{\varepsilon}(s), s\le t\}$. We say $\text{p-lim}_{\delta\to 0} f^{\delta} = 0$ iff $\sup_{\delta,t} E|f^{\delta}(t)| < \infty$ and $E|f^{\delta}(t)| \to 0$ as $\delta \to 0$. Define an operator $\hat A^{\varepsilon,N}$ and its domain $\mathscr{D}(\hat A^{\varepsilon,N})$ as follows: $g \in \mathscr{D}(\hat A^{\varepsilon,N})$ and $\hat A^{\varepsilon,N}g = q$ iff $g,q \in \mathscr{M}$ and

$$\underset{\delta\downarrow 0}{\text{p-lim}}\; E\left|\frac{E_t^{\varepsilon,N}g(t+\delta)-g(t)}{\delta} - q(t)\right| = 0.$$

The following theorem is Theorem 1 of [2], adapted to our case. $\hat{\mathscr{C}}_0$ denotes the set of continuous real-valued functions on $R^r \times [0,\infty)$.

Theorem 1. Let the equation (1.2) have a unique weak-sense solution. Fix N. For each $f(\cdot) \in \mathcal{D}$, a dense set (sup norm) in $\hat{\mathcal{L}}_0$, let there be a sequence $\{f^{\varepsilon,N}(\cdot)\} \in \mathcal{A}$ satisfying the following:

$$\text{(3.2)} \qquad \underset{\substack{\varepsilon\to 0\\ \alpha\to 0}}{\text{p-lim}} \, |f^{\varepsilon,N}(t) - f(v^{\varepsilon,N}(t),t)| \to 0,$$

$$\text{(3.3)} \qquad f^{\varepsilon,N}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon,N}),$$

$$\text{(3.4)} \qquad \underset{\substack{\varepsilon\to 0\\ \alpha\to 0}}{\text{p-lim}} \, |\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(t) - (A^N + \tfrac{\partial}{\partial t}) f(v^{\varepsilon,N}(t),t)| = 0,$$

where A^N is the infinitesimal operator of some diffusion process and the coefficients of A^N and A are equal for $v \in S_N$. Then if $\{v^{\varepsilon,N}(\cdot)\}$ is tight for each N, $\{v^{\varepsilon}(\cdot)\} \to v(\cdot)$ weakly.

Comment. Tightness is not hard to prove here. See comments at the end of the proof of Theorem 2, which applies Theorem 1 to our case (1.1). Given $f(\cdot,\cdot)$, the main problem is to find the $f^{\varepsilon,N}(\cdot)$ and to verify (3.2) - (3.4) (and ultimately to prove tightness). The method used here and in [1], [2] is similar to the averaging method used in [3]. We choose the form $f^{\varepsilon,N}(t) = f(v^{\varepsilon,N}(t),t) + f_1^{\varepsilon,N}(t) + f_2^{\varepsilon,N}(t)$, where $f_1^{\varepsilon,N}(t)$ is chosen so as to "average out" certain noise-dependent terms in $\hat{A}^{\varepsilon,N} f(v^{\varepsilon,N}(t),t)$, and $f_2^{\varepsilon,N}(t)$ is chosen to "average out" certain noise-dependent terms which result from applying $\hat{A}^{\varepsilon,N}$ to $f_1^{\varepsilon,N}(t)$. In the proof $\lim_{\substack{\varepsilon\to 0\\ \alpha\to 0}}$ means $\lim_{\varepsilon\to 0} \lim_{\alpha\to 0}$.

IV. The Convergence Theorem

Theorem 2. Under the assumptions in Section II, $\{v^{\varepsilon}(\cdot)\}$ converges weakly to $v(\cdot)$ as $\varepsilon \to 0$ and then $\alpha \to 0$.

Proof. Let $\mathscr{D} = \hat{\mathscr{C}}_0^{2,3}$, the subspace of $\hat{\mathscr{C}}_0$ of functions whose mixed partial derivatives up to order 2 in t and 3 in x are continuous. By Theorem 1, for each N and $f(\cdot,\cdot) \in \mathscr{D}$, we only need to find $\{f^{\varepsilon,N}(\cdot)\}$ satisfying (3.2) - (3.4). For notational convenience write $v^{\varepsilon,N}(\cdot)$ as $v^{\varepsilon}(\cdot)$ in this proof, but we are always working with the truncated process $v^{\varepsilon,N}(\cdot)$.

Part 1. Fix $f(\cdot,\cdot) \in \mathscr{D}$. Then

$$(4.1) \quad \hat{A}^{\varepsilon,N} f(v^{\varepsilon}(t),t) = f_t(v^{\varepsilon}(t),t) +$$

$$+ b_N(v^{\varepsilon}(t)) f_v'(v^{\varepsilon}(t),t) \cdot [F(v^{\varepsilon}(t)) + \frac{DL}{\varepsilon} g_{\alpha}(s(t)+n^{\varepsilon}(t)-G(v^{\varepsilon}(t)))].$$

Note that for u in any bounded set

$$(4.2) \quad \frac{1}{\varepsilon} E g_{\alpha}(u+n^{\varepsilon}(t)) \equiv \hat{g}_{\alpha}^{\varepsilon}(u)/\varepsilon = \frac{1}{\varepsilon}[P\{z(0)>-\varepsilon u+\varepsilon\alpha\}-P\{z(0)<-\varepsilon u-\varepsilon\alpha\} + O(\varepsilon\alpha)/\varepsilon]$$

$$= \sqrt{2/\pi}\ \frac{u}{\sigma} + O(\alpha) + O(\varepsilon),$$

which justifies the $L_{\varepsilon} = L/\varepsilon$ scaling. We will get $f^{\varepsilon,N}(\cdot)$ in the form

$$f^{\varepsilon,N}(t) = f(v^{\varepsilon}(t),t) + f_1^{\varepsilon,N}(t) + f_2^{\varepsilon,N}(t),$$

where the $f_i^{\varepsilon,N}(\cdot)$ will be defined below.

The following estimate will be used.

(4.3) On the set $\{|z(t)|\le 1\}$ or even on $\{|z(t)|\le e^{a\tau/2}\}$, $|P\{z(t+\tau)\in B|z(t)\} - P\{z(t+\tau)\in B\}| \le Ce^{-a_1\tau}$ for some constants C and $a_1 > 0$, uniformly in B. Similarly, on the same $z(t)$ set and for $\tau_2 > \tau_1 > 0$,
$|P\{z(t+\tau_i)\in B_i,\ i=1,2|z(t)\} - P\{z(\tau_i+t)\in B_i,\ i=1,2\}| \le Ce^{-a_1\tau_1}$ for some $a_1 > 0$ and $C < \infty$ and all B_1, B_2.

In the sequel the values of a_1 and C may change from usage to usage. Define $\bar{g}_\alpha(u,n^\varepsilon(t)) = g_\alpha(u+n^\varepsilon(t)) - Eg_\alpha(u+n^\varepsilon(t))$ and define

$$
\begin{aligned}
f_1^{\varepsilon,N}(t) &= \frac{L}{\varepsilon} b_N(v^\varepsilon(t))\int_0^\infty f_v'(v^\varepsilon(t),t+\tau)DE_t^{\varepsilon,N}\bar{g}_\alpha(s(t+\tau)-G(v^\varepsilon(t)),n^\varepsilon(t+\tau))d\tau \\
&= \varepsilon L b_N(v^\varepsilon(t))\int_0^\infty f_v'(v^\varepsilon(t),t+\varepsilon^2\tau)DE_t^{\varepsilon,N}\bar{g}_\alpha(s(t+\varepsilon^2\tau)-G(v^\varepsilon(t)),z(\frac{t}{\varepsilon^2}+\tau)/\varepsilon)d\tau
\end{aligned}
$$

By (4.3), $f_1^{\varepsilon,N}(t) = O(\varepsilon)$ uniformly in ω, α, on the set $\{|z(t/\varepsilon^2)|\le 1\}$. Define $w_1 = \min\{\tau\colon e^{-a\tau/2}|z(t/\varepsilon^2)|\le 1\}$. Write $f_1^{\varepsilon,N}(\cdot)$ as

$$
f_1^{\varepsilon,N}(t) = \varepsilon\int_0^{w_1} E_t^{\varepsilon,N}\{\cdot\}\,d\tau + \varepsilon\int_{w_1}^\infty E_t^{\varepsilon,N}\{\cdot\}\,d\tau.
$$

The first term is bounded in absolute value by $\varepsilon C w_1$ and the integrand of the second by $C\exp -a_1\tau$. Thus

(4.4) $|f_1^{\varepsilon,N}(t)| \le C\varepsilon(1+w_1) \le C\varepsilon[1 + \max(0,\log|z^\varepsilon(t)|)]$.

Part 2. It can be verified that $f_1^{\varepsilon,N}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon,N})$ and that

$$\text{(4.5)} \quad \hat{A}^{\varepsilon,N} f_1^{\varepsilon,N}(t) = -\frac{Lb_N(v^\varepsilon(t))}{\varepsilon} f_v'(v^\varepsilon(t),t) D\bar{g}_\alpha(s(t)-G(v^\varepsilon(t)),n^\varepsilon(t))$$

$$+\frac{L}{\varepsilon}\int_0 E_t^{\varepsilon,N}[b_N(v^\varepsilon(t)) f_v'(v^\varepsilon(t),t+\tau) D\bar{g}_\alpha(s(t+\tau)-G(v^\varepsilon(t)),n^\varepsilon(t+\tau))]_v' \dot{v}^\varepsilon(t),$$

where the subscript* v denotes the gradient of the bracketed expression with respect to $v^\varepsilon(t)$. At this point, let us simplify the notation by dropping the $b_N(v)$ terms. All of the $f_i^{\varepsilon,N}$ will be proportional to either $b_N(v^\varepsilon(t))$ or $b_N^2(v^\varepsilon(t))$. Changing variables $\tau/\varepsilon^2 \to \tau$ and splitting the integral in (4.5) into two parts and using $\dot{v}^\varepsilon = b_N(v^\varepsilon)[s-G(v^\varepsilon)+Dg_\alpha/\varepsilon]$ (but dropping the $b_N(v)$) yields

$$\text{(4.6)} \quad L^2\int_0^\infty E_t^{\varepsilon,N}(D'f_v(v^\varepsilon(t),t+\varepsilon^2\tau))_v' Dg_\alpha(s(t+\varepsilon^2\tau)-G(v^\varepsilon(t)),z(\tfrac{t}{\varepsilon^2}+\tau)/\varepsilon)\,\cdot$$

$$g_\alpha(s(t)-G(v^\varepsilon(t))+z(t/\varepsilon^2)/\varepsilon)d\tau + O(\varepsilon)$$

$$+L^2\int_0^\infty E_t^{\varepsilon,N}(D'f_v(v^\varepsilon(t),t+\varepsilon^2\tau))\bar{g}_{\alpha,v}'(s(t+\varepsilon^2\tau)-G(v^\varepsilon(t),z(\tfrac{t}{\varepsilon^2}+\tau)/\varepsilon)D\,\cdot$$

$$g_\alpha(s(t)-G(v^\varepsilon(t))+z(t/\varepsilon^2)/\varepsilon)d\tau + O(\varepsilon).$$

The terms in (4.6) exist by the same arguments which led to (4.4). We next show that the second integral of (4.6) is negligible as $\varepsilon \to 0$, $\alpha \to 0$ and get an estimate which is useful for the tightness argument. The facts that $s(t)$ and $v^\varepsilon(t)$ are bounded (recall that we are using the truncated process (3.1)) and that the support of $g_{\alpha,u}(u)$ is* in $[-\alpha,\alpha]$ and that $|g_{\alpha,u}(u)| \le C/\alpha$

---

*$g_{\alpha,u}(\cdot)$ is the derivative of $g_\alpha(\cdot)$ with respect to its argument. The subscript v denotes the derivative with respect to the explicit argument v: replace $v^\varepsilon(t)$ by v, take the derivative with respect to v, then set $v = v^\varepsilon(t)$.

will be used frequently and perhaps without specific mention. Let $I\{A\}$ denote the indicator function of the set A.

By (4.3) it can be verified that

$$(4.7) \qquad Y \equiv |E_t^{\varepsilon,N}\bar{g}_{\alpha,v}(s-G(v),z(\frac{t}{\varepsilon^2}+\tau)/\varepsilon)| \leq [\exp -a_1\tau + I\{|z(t/\varepsilon^2)|>e^{a\tau/2}\}]C/\varepsilon.$$

We need a bound on Y which goes to zero as $\varepsilon \to 0$. First we get such a bound when $|z(t/\varepsilon^2)| \geq 1$. Note that

$$(4.8) \qquad P\{|s-G(v)+z(\frac{t}{\varepsilon^2}+\tau)/\varepsilon| \leq \alpha \Big| z(t/\varepsilon^2)=z_0\} = O(\alpha\varepsilon)$$

uniformly for $|z_0| \geq 1$ and $\tau > 0$ (recall that s, v are in a bounded set - for each N). Now, (4.8) and the facts cited above (4.7) imply that Y is bounded by $O(\varepsilon)$, uniformly in $|z(t/\varepsilon^2)| \geq 1$, $\tau \geq 0$. Thus on $|z(t/\varepsilon^2)| \geq 1$,

$$(4.9) \qquad Y \leq [\exp -a_1\tau + I\{|z(t/\varepsilon^2)|>e^{a\tau/2}\}]^{1/2}C(\varepsilon/\alpha)^{1/2}$$

(use $|x| \leq a$, $|x| \leq b \Rightarrow |x| \leq \sqrt{ab}$). Thus, on integrating the bound when $|z(t/\varepsilon^2)| \geq 1$, we see that the second term of (4.6) is bounded above by

$$C[1 + \max(0, \log|z(t/\varepsilon^2)|)] \ (\varepsilon/\alpha)^{1/2}.$$

Now, we look for a bound when $|z(t/\varepsilon^2)| \leq 1$. Split the second integral in (4.6) into the two parts $\int_0^\varepsilon + \int_\varepsilon^\infty$. The first part is $O(\varepsilon/\alpha)$. Note that the density of $z(\frac{t}{\varepsilon^2}+\tau)$, $\tau \geq \varepsilon$, conditioned on any value of $|z(t/\varepsilon^2)|$ in [0,1], is bounded above by $O(1/\sqrt{\varepsilon})$. So (4.8) then holds with $O(\varepsilon\alpha)$ replaced by $O(\varepsilon^{1/2}\alpha)$.

Combining this estimate with (4.7) yields that Y is bounded above by (4.9) when $|z(t/\varepsilon^2)| \le 1$, but with the change that $(\varepsilon/\alpha)^{1/2}$ is replaced by $(\varepsilon^{1/4}/\alpha^{1/2})$ in (4.9). Thus, on integrating the bound, we get that the second term of (4.6) is bounded above by

$$(4.10)\qquad C[1+\max(0,\ \log|z(t/\varepsilon^2)|)]\varepsilon^{1/4}/\alpha^{1/2}.$$

Part 3. We turn our attention to the first term of (4.6) and show that, by an "averaging", it can effectively be replaced by its expectation. To facilitate the development, we define the following terms.

$$h_\varepsilon(v,t,\tau,\rho) \equiv L^2 D' f_{vv}(v,t+\tau+\rho)D\cdot\bar{g}_\alpha(s(t+\tau+\rho)-G(v)+n^\varepsilon(t+\tau+\rho))\ \cdot$$

$$\cdot\ g_\alpha(s(t+\rho)-G(v)+n^\varepsilon(t+\rho)),$$

$$H_\varepsilon(v,t,\tau,\rho) \equiv L^2 D' f_{vv}(v,t+\varepsilon^2\tau+\varepsilon^2\rho)D\cdot\bar{g}_\alpha(s+(t+\varepsilon^2\tau+\varepsilon^2\rho)-G(v)+z(\tfrac{t}{\varepsilon^2}+\tau+\rho)/\varepsilon)\ \cdot$$

$$\cdot\ g_\alpha(s(t+\varepsilon^2\rho)-G(v)+z(\tfrac{t}{\varepsilon^2}+\rho)/\varepsilon),$$

$$(4.11)\qquad A_0^{\varepsilon,N}f(v,t) \equiv \int_0^\infty EH_\varepsilon(v,t,\tau,0)\,d\tau = \frac{1}{\varepsilon^2}\int_0^\infty Eh_\varepsilon(v,t,\tau,0)\,d\tau.$$

$$(4.12)\qquad f_2^{\varepsilon,N}(t) \equiv \frac{1}{\varepsilon^2}\int_0^\infty d\rho \int_0^\infty d\tau\,[E_t^{\varepsilon,N}h_\varepsilon(v^\varepsilon(t),t,\tau,\rho) - Eh_\varepsilon(v,t,\tau,\rho)\Big|_{v=v^\varepsilon(t)}]$$

$$= \varepsilon^2\int_0^\infty d\rho \int_0^\infty d\tau\,[E_t^{\varepsilon,N}H_\varepsilon(v^\varepsilon(t),t,\tau,\rho) - EH_\varepsilon(v,t,\tau,\rho)\Big|_{v=v^\varepsilon(t)}],$$

where $\Big|_{v\ v^{\varepsilon}(t)}$ implies that v is replaced by $v^{\varepsilon}(t)$ after taking the expectation. We must show that $f_2^{\varepsilon,N}(\cdot)$ is well defined. First note that the inner integral of $f_2^{\varepsilon,N}(t)$ (with $\rho = 0$ and a change of variables) is just the first term of (4.6) centered about its expectation. The form of $f_2^{\varepsilon,N}(\cdot)$ is chosen to allow us to average out the first term of (4.6) and to effectively replace it by its <u>average value</u> $A_0^{\varepsilon,N}f(v^{\varepsilon}(t),t)$.

By the method used to bound $|f_1^{\varepsilon,N}(t)|$, we can get that the inner integral of $f_2^{\varepsilon,N}(t)$ exists for each ρ. Recall the definition $w_1 = \min\{w: e^{-aw/2}|z(t/\varepsilon^2)|\leq 1\}$ and write (4.12) as

$$\varepsilon^2\int_{w_1}^{\infty} d\rho \int_0^{\infty} d\tau\ E_t^{\varepsilon,N}B + \varepsilon^2\int_0^{w_1} d\rho \int_0^{\infty} d\tau\ E_t^{\varepsilon,N}B \quad \mathrm{II} + \mathrm{I}.$$

First we show that II is well defined. By (4.3) and the definition of w_1, the absolute value of the integrand in II is bounded above by $C \exp -a_1\rho$, $a_1 > 0$. Also $|EH_{\varepsilon}(v,t,\tau,\rho)| \leq C \exp -a_1\tau$ for some $a_1 > 0$. By (4.3) <u>and on the set</u> $\{\rho\geq w_1\}$, and for C, a_1 (whose values again may change from usage to usage)

$$\begin{aligned}
|E_t^{\varepsilon,N}H_{\varepsilon}(v,t,\tau,\rho)| &\leq E_t^{\varepsilon,N}|E_{t+\varepsilon^2\rho}^{\varepsilon,N}H_{\varepsilon}(v,t,\tau,\rho)| \\
&\leq C\ E_t^{\varepsilon,N}[\exp -a_1\tau + I\{e^{-a\tau/2}|z(\tfrac{t}{\varepsilon^2}+\rho)|\geq 1\}] \\
&= C \exp -a_1\tau + CP\{|z(\tfrac{t}{\varepsilon^2}+\rho)|>e^{a\tau/2}\big|\ \rho\geq w_1; z(t/\varepsilon^2)\} \\
&\leq C \exp -a_1\tau + Ce^{-a\tau/2}E\{|z(\tfrac{t}{\varepsilon^2}+\rho)|\big|\ \rho\geq w_1; z(t/\varepsilon^2)\} \\
&\leq C \exp -a_1\tau.
\end{aligned}$$

Chebychev's inequality is used to get the next-to-last inequality. Combining the above estimates yields that the integrand in II is bounded by (for some $a_1 > 0$, $C < \infty$) $C \exp -a_1(\tau+\rho)$. Hence $II = O(\varepsilon^2)$.

The term I is also $O(\varepsilon^2)$ but not uniformly in $z(t/\varepsilon^2)$. Bound the inner integral of I by

$$\int_0^\infty d\tau \left|E_t^{\varepsilon,N} B\right| \leq E_t^{\varepsilon,N} \int_0^\infty d\tau \ \left|E^{\varepsilon,N}_{t+\varepsilon^2\rho} B\right| \quad III.$$

By the arguments used to get the bound on $|f_1^{\varepsilon,N}(t)|$, we get

$$III \leq C\varepsilon^2 \ E_t^{\varepsilon,N}[1 + \max(0, \ \log|z(\frac{t}{\varepsilon^2}+\rho)|)]$$

$$\leq C\varepsilon^2 \ E_t^{\varepsilon,N}[1 + \log(|z(\frac{t}{\varepsilon^2}+\rho)| + 1)].$$

By Jensen's inequality and the concavity of $\log(\cdot)$,

$$III \leq C\varepsilon^2[1 + \log(|z(\frac{t}{\varepsilon^2})| + C)].$$

Since $w_1 \leq C \max(0, \ \log|z(t/\varepsilon^2)|)$,

$$(4.13) \quad |f_2^{\varepsilon,N}(t)| \leq C\varepsilon^2[1 + \log(|z(\frac{t}{\varepsilon^2})| + C)]^2.$$

Henceforth, we will give only an outline of the details, which can all be filled in via the estimates and techniques developed above. It can be shown that $f_2^{\varepsilon,N}(\cdot) \in \mathscr{D}(\hat{A}^{\varepsilon,N})$ and that

$$(4.14)\quad \hat{A}^{\varepsilon,N} f_2^{\varepsilon,N}(t) = \text{negative of first term of (4.6)} + A_0^{\varepsilon,N} f(v^{\varepsilon}(t),t)$$

$$+ \text{(terms whose } \underset{\substack{\varepsilon\to 0\\ \alpha\to 0}}{\text{p-lim}} \text{ equal zero)}.$$

The term whose p-lim = 0 is just $(f_{2,v}^{\varepsilon,N}(t))'\dot{v}^{\varepsilon}(t)$, where $f_{2,v}^{\varepsilon,N}$ is the gradient of the expression for $f_2^{\varepsilon,N}$ with respect to the argument $v^{\varepsilon}(t)$. The components of $\hat{A}^{\varepsilon,N} f_2^{\varepsilon,N}$ which involve f_{vvv} are bounded by $O(\varepsilon)$. Loosely speaking, the remaining component is of the form

$$(4.15)\quad o(\varepsilon) + \varepsilon\iint [E_t^{\varepsilon,N} f_{vv}\bar{g}_{\alpha,v} g_{\alpha} + E_t^{\varepsilon,N} f_{vv}\bar{g}_{\alpha} g_{\alpha,v} - E\, f_{vv}\bar{g}_{\alpha,v} g_{\alpha}$$

$$- E\, f_{vv}\, \bar{g}_{\alpha} g_{\alpha,v}]d\tau d\rho \cdot g_{\alpha},$$

where we omit the function arguments. By a method similar to that used to get (4.13), we get the bound (4.13) on (4.15) but with (ε/α) replacing ε^2.

Part 4. The estimates obtained in Parts 1 - 3 imply that

$$\underset{\substack{\varepsilon\to 0\\ \alpha\to 0}}{\text{p-lim}} \left|f^{\varepsilon}(t) - f(v^{\varepsilon}(t),t)\right| = 0,$$

$$\underset{\substack{\varepsilon\to 0\\ \alpha\to 0}}{\text{p-lim}} \Big|\hat{A}^{\varepsilon,N} f^{\varepsilon}(t) - f_t(v^{\varepsilon}(t),t) - L\sqrt{2/\pi}\,\frac{(s(t)-G(v^{\varepsilon}(t)))}{\sigma}\, D' f_v(v^{\varepsilon}(t),t)$$

$$- A_0^{\varepsilon,N} f(v^{\varepsilon}(t),t)\Big| = 0.$$

A proof very similar to that in ([5], Section 6, part 2) yields* that $A_0^{\varepsilon,N} f(v,t) \to D' f_{vv}(v,t) D(\ln 2)/a$ uniformly in v for each t. In calculating the limit, the

*One of the reasons for the choice $\text{cov}[z(0),z(t)] = \sigma^2 \exp -\alpha\tau$ is to allow us to save work by using this result. The choice allowed an explicit evaluation of the diffusion term. With other choices the diffusion coefficient would be left in an "integral" form.

$G(\cdot)$, $s(\cdot)$ play no role and the limit $(\varepsilon \to 0, \alpha \to 0)$ is the same as for the case $(\alpha = 0, \varepsilon \to 0)$. If the $b_N(v)$ terms were retained, the result would be the same, except that either b_N or b_N^2 would multiply the f_v, f_{vv} or f_{vvv}. By what has just been said

$$\text{(4.16)} \qquad \underset{\substack{\varepsilon\to 0\\ \alpha\to 0}}{\text{p-lim}}\ \left|\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(t) - \left(\frac{\partial}{\partial t} + A\right) f(v^{\varepsilon}(t),t)\right| = 0,$$

where A is the infinitesimal operator of $v(\cdot)$ in (1.2). If b_N were retained, the A in (4.16) would be replaced by some A^N which would equal A where $b_N(\cdot) = 1$, i.e. in S_N. Thus, by Theorem 1, if $\{v^{\varepsilon,N}(\cdot)\}$ were tight, then the proof would be completed.

<u>Tightness</u>. Use ([2], Theorem 2). The conditions of Theorem 2 [2] hold if (4.17) holds for each N and $T < \infty$:

$$\text{(4.17)} \qquad \lim_{K\to\infty}\ \overline{\lim_{\substack{\varepsilon\to 0\\ \alpha\to 0}}}\ P\{\sup_{t\le T}|\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(t)| \ge K\} = 0,$$

$$\lim_{\substack{\varepsilon\to 0\\ \alpha\to 0}} P\{\sup_{t\le T} |f_1^{\varepsilon,N}(t) + f_2^{\varepsilon,N}(t)| \ge \delta\} = 0, \qquad \text{each } \delta > 0.$$

But (4.17) follows from (4.10), (4.13) (and a similar estimate for (4.15), and the fact that the Gaussianness and stationarity imply that for any $\gamma > 0$

$$\lim_{\varepsilon\to 0}\ \sup_{t\le T} \varepsilon^{\gamma}|z(t/\varepsilon^2)| = 0 \qquad \text{w.p. 1.} \qquad\qquad \text{Q.E.D.}$$

References

1. H.J. Kushner, "Jump-diffusion approximations for ordinary differential equations with random right-hand sides", SIAM J. on Control and Optimization, 17, 729-744, 1979.

2. H.J. Kushner, "A martingale method for the convergence of a sequence of processes to a jump-diffusion process", to appear Z. Wahrscheinlichkeitsteorie.

3. G. Blankenship, G.C. Papanicolaou, "Stability and control of stochastic systems with wide-band noise disturbances", SIAM J. on Applied Math., 34, 437-476, 1978.

4. G.C. Papanicolaou, W. Kohler, "Asymptotic theory of mixing ordinary differential equations", Comm. Pure and Appl. Math., 27, 641-668, 1974.

5. H.J. Kushner, "Diffusion approximations to output processes of nonlinear systems with wide-band inputs and applications to communication theory", to appear IEEE Trans. on Information Theory.

6. P. Billingsley, Convergence of probability measures, John Wiley & Sons, New York, 1968.

7. T.G. Kurtz, "Semigroups of conditional shifts and approximation of Markov processes", Ann. Prob., 4, 618-642, 1975.

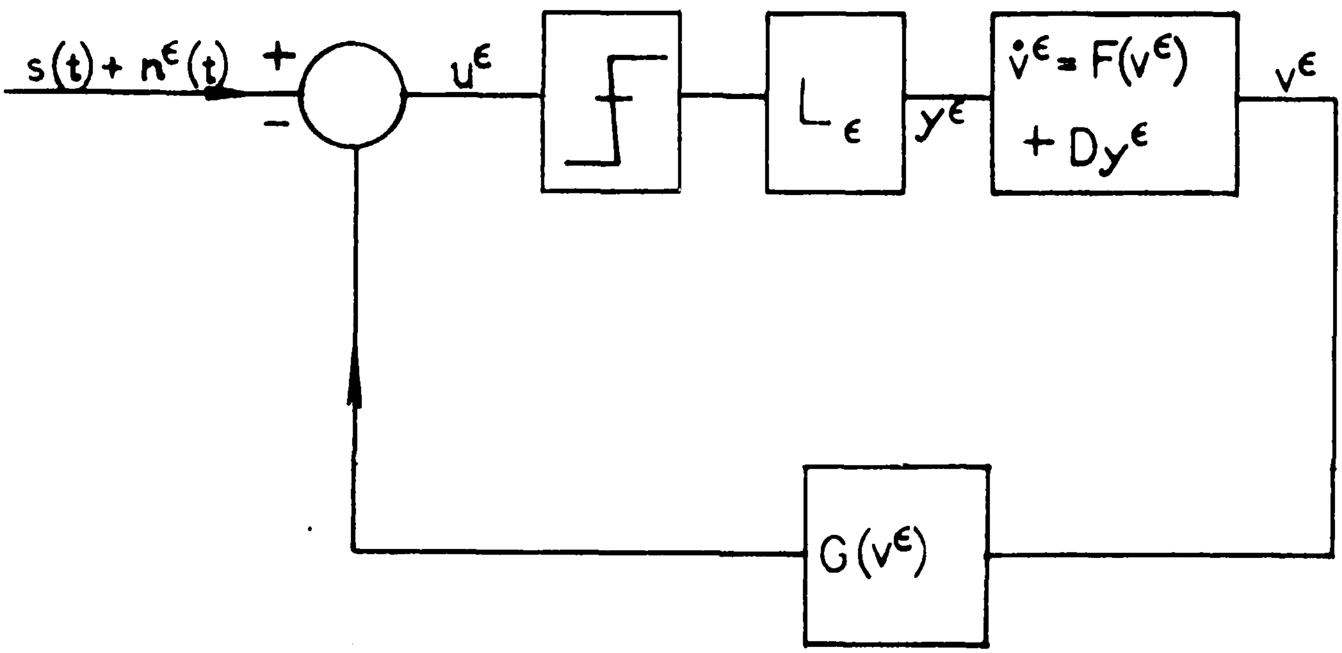


FIG.1. THE NONLINEAR SYSTEM
WITH A SATURATOR

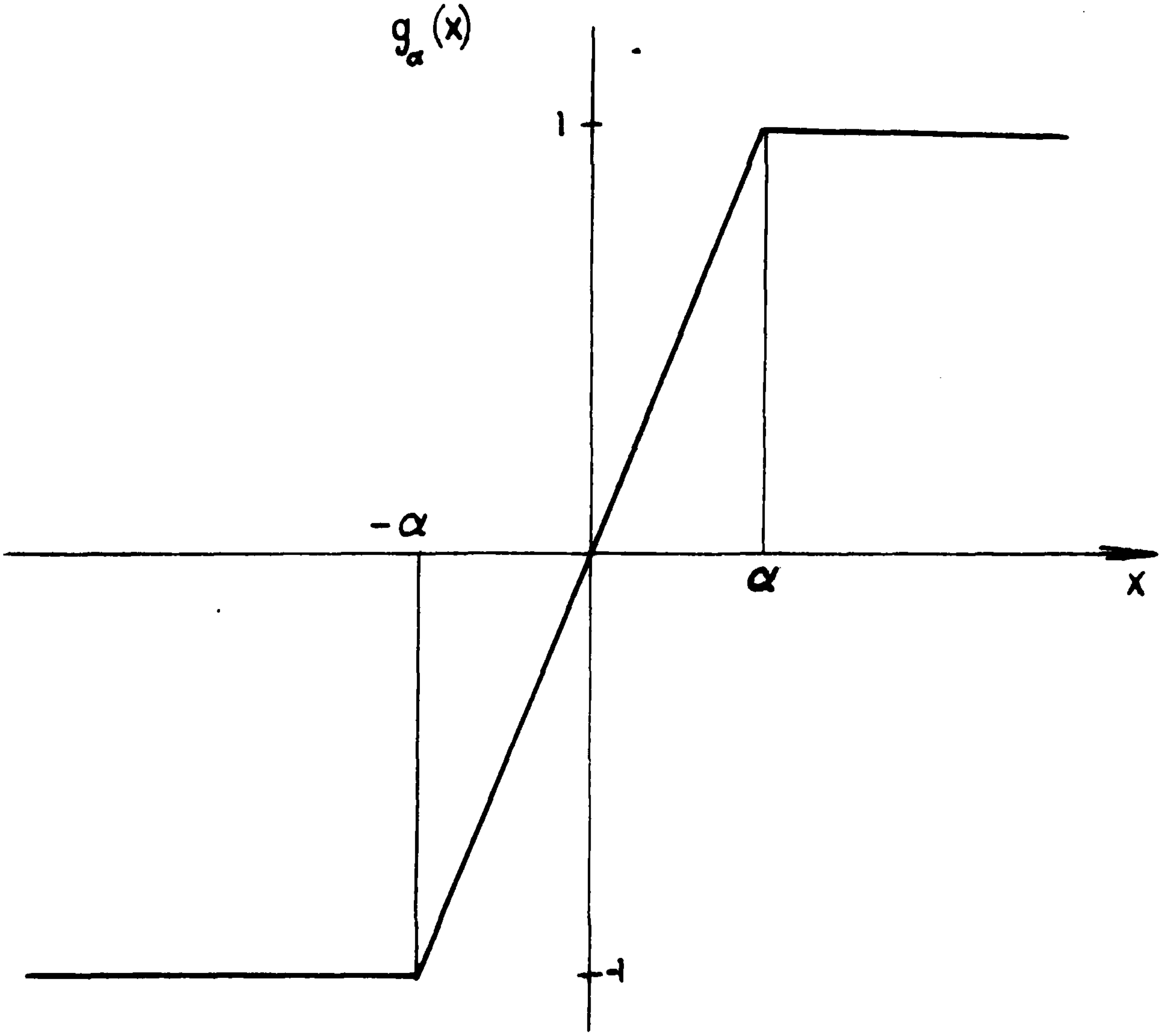


FIG.2. THE APPROXIMATE SATURATOR


9